Surgical Oncology 20 (2011) 169–177

Contents lists available at ScienceDirect

Surgical Oncology

journal homepage: www.elsevier.com/locate/suronc

Review

Non-technical skills assessment in surgery

Bharat Sharma [a,*], Amit Mishra [b], Rajesh Aggarwal [b], Teodor P. Grantcharov [a]

[a] Department of Surgery, University of Toronto, Division of General Surgery, St. Michael's Hospital, Toronto, Ontario, Canada
[b] Department of Biosurgery and Surgical Technology, Imperial College, London, UK

Keywords:
Non-technical skills
Human factors
NOTSS
NOTECHS
OTAS
ANTS
Decision making
Communication
Leadership
Managerial skills
Situation awareness

A B S T R A C T

Adverse events in surgery have highlighted the importance of non-technical skills, such as communication, decision-making, teamwork, situational awareness and leadership, to effective organizational performance. These skills carry particular importance to surgical oncology, as members of a multidisciplinary team must work cohesively to formulate effective patient care plans. Several non-technical skills evaluation tools have been developed for use in surgery, without adequate comparison and consensus on which should be standard for training.

Eleven articles describing the use of three non-technical evaluation tools related to surgery: NOTSS (Non Technical Skills for Surgeons), NOTECHS (Non Technical Skills) and OTAS (Observational Teamwork Assessment for Surgery) were analyzed with respect to scale formulation, validity, reliability and feasibility. Furthermore, their use in training thus far and the future of non-technical rating scales in surgical curricula was discussed. Future work should focus on incorporating these assessment tools into training and into a real operating room setting to provide formative evaluations for surgical residents.

Contents

* Corresponding author. Tel.:+1 647 281 3959.
E-mail address: b.sharma@utoronto.ca (B. Sharma).

0960-7404/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.suronc.2010.10.001

Introduction

Industries such as aviation, nuclear engineering, energy exploration and medicine can be considered high risk, as errors associated with them may be devastating, both in terms of outcome and the inevitable compromise to public perceptions of the organizations involved. The knowledge base and technical competency of these industry workers is essential, and organizations invest significant resources to ensure their technical competence. However factors such as communication, leadership, resource management and decision making, termed together as 'non-technical skills' (NTS) are equally important with analysis of adverse events revealing that as individual technical competence plateaus at high levels it is increasingly deficiencies in NTS that lead to adverse events in increasingly complex organizational structures. Specifically over the last 20 years, the importance of NTS for delivering safe and high-quality medical care has been increasingly recognized [1–5]. Much of the work that exists now stems from the aviation industry, where NTS have been empirically identified and trained through Crew Resource Management (CRM) courses [6]. In surgical oncology, these skills are particularly important, as medical oncologists, surgeons, pathologists, along with other specialists, all with varied training backgrounds and differing paradigms on ideal patient care must work together to devise care plans.

Within the medical field behavioural marker systems that evaluate non-technical skills are already in use. These marker systems are rating scales based on skills taxonomies and are used to evaluate observable behaviours that underlie performance [7]. In anaesthesia, a taxonomy of NTS that is important for safety in the operating theatre, the Anaesthetists' Non-Technical Skills (ANTS) behavioural marker system has been developed [8]. Similarly, tools have been developed to assess NTS in surgical trainees. This shift in focus comes from analyses of adverse events during surgery that point to failings in NTS rather than technical expertise alone, as causative factors [9–11].

As with any assessment system, these behavioural rating tools must be explicit, transparent, valid and reliable for effective assessment of NTS [12,13]. The purpose of the present study was to review the literature on non-technical skills rating scales used in surgery. Furthermore, the aim was to provide a background; define the specific elements of each tool, comment on their reliability, validity and feasibility as tools that can be incorporated into future surgical curricula.

Methods

Search method

On-line sources including BioMed Central, PubMed Medline, Cochrane Library, the ERIC Education database, EMBASE plus bibliographies from related research papers were consulted. The search terms used were: Surgery/General Surgery/Urology/Cardiac Surgery/Medicine/Internal Medicine/Critical Care AND Non-technical/education/Non-technical skills/Team Performance and the search was limited to the English language and limited to articles published from 1959 to August 2010.

Search outcome

The search revealed a total of 178 publications, which were subsequently screened to ensure that they were relevant to the area of interest for this study. All of these were published in peer-reviewed journals and none of them were review articles. Of the original papers, only 11 were found to be relevant to assessment of non-technical skills of surgical residents. These papers were systematically reviewed to identify evaluation tools, training programs and the context in which these non-technical skill programs were formulated. The literature was analyzed to understand the validity, reliability and clinical application of the current rating scales.

Results

Two major skills taxonomies, namely the NOTSS [14,15] and NOTECHS [16] systems have been developed over the last decade to evaluate non-technical skills in surgery (Table 1). Both are grounded in studies demonstrating non-technical skill errors in the operating room environment and from lessons learnt from evaluation tools in high risk industries, such as aviation and nuclear power, highlighting the importance of non-technical skills. Additionally, a teamwork assessment tool in surgery (OTAS) and non-technical skills assessment tools in non surgical domains have also been developed.

Requirements for an assessment tool of non-technical skills

All behavioural-rating systems must be evaluated according to standardized psychometric criteria [12,13]. Specifically, they must be valid and therefore capture what they purport to capture. Second, they must be reliable, which indicates good test–retest correlation and consistent results across various administrations. Third, they must be sensitive and detect differences within elements when they actually exist. Additionally, an assessment tool must be feasible, which includes cost effectiveness, easy to implement (without significant hindrance to clinical practice) and be easy to adopt into real time/simulated environments without significant time/resource demands. Furthermore, it should provide meaningful feedback and be transparent, in that each it should describe each aspect of the evaluation and give descriptors that indicate good vs poor performance.

Non Technical Skills for Surgeons (NOTSS)

Scale development

The Non Technical Skills for Surgeons (NOTSS) rating scale was developed with a focus on surgeons' intra-operative non-technical skills [15]. It was developed in 2006 by interviewing 27 consultant surgeons in various surgical specialties, using the critical incident technique, a type of cognitive interview. Several unique questioning

Table 1
Summary of non-technical rating scales.

Article	Scale	Population Studied	Methodology	Reliability	Validity	Comments
NOTECHS						
Sevdalis et al. [16]	Revised NOTECHS	OR Team	Non-randomized experimental study. Simulated OR crisis scenarios with and without non-technical skills training intervention	Good reliability across raters (Cronbach's alpha 0.77 to 0.87) Good reliability across expert trainers and trainees (Cronbach's alpha 0.76 to 0.89) Pre and post intervention reliability (Cronbach's alpha 0.85 to 0.9)	N/A	Good for team and individual non-technical skills performance rating Needs further validity testing, which at present is grounded in aviation literature Not evaluated in real time operating room environment
Mishra et al. [29]	Oxford NOTECHS system, OCHRA, SAQ, OTAS	Overall Team and each subteam	Non randomized observational study. Pre and post non-technical skill training performance evaluated.	Good reliability across raters (Cronbach's alpha 0.95–0.99) Scoring was similar pre and post training across raters.	Restricted to OR non crisis scenario – as this was the framework of present study Improved Oxford NOTECHS scores after training intervention and correlate with attitude surveys to teamwork, with OTAS and with surgical errors during cases..	Team oriented Scale with specific behavioural descriptors for each OR subteam (e.g. nursing, surgical) Used Expert Raters – with prior experience in non-technical skills assessment Correlated non-technical skills performance with technical errors and another non-technical rating scale.
Mishra et al. [27]	NOTECHS, OCHRA	Overall Team Performance and each member of the OR team	Non randomized observational study. Real OR cases (Lap. Cholecystectomy) rated by 2 raters.	Good inter-rater reliability (Cronbach's alpha 0.88) for the observed cases	N/A	Used to rate overall team and individual subteam performance in real OR environment. Used experienced raters for evaluation of cases
McCulloch et al.[28]	Oxford NOTECHS SAQ, OCHRA and OTE's also assessed	OR Team and each subteam scored separately	Non randomized observational real time study. OR cases (Lap. Cholecystectomy andCarotid enderarterectomy)	Good inter-rater reliability across all cases observed (Cronbach's alpha 0.98)	Improved Oxford NOTECHS scores after training intervention and correlated with positive attitude surveys of teamwork. Nontechnical skills training resulted in fewer procedural and non procedural errors	9 h. Non technical skills training course followed by CRM coaching for 3 months Demonstrates association of Team Nontechnical skills performance with technical errors.
NOTSS						
Yule et al. [20]	NOTSS	Pre-Recorded Surgeon Behaviours – to test the NOTSS rating system	Experimental study – with simulated operating room scenarios demonstrating a range of good to poor behaviours	Variability existed between novice (minimally trained) and expert raters on scoring the scenarios. Ratings were more reliable in the face of extreme behaviours (i.e. either excellent or very poor behaviours)	NOTSS developed using a bottom up approach with subject matter experts. Task analysis, iterative process, revision after psychometric evaluation	Focuses on a surgeon's non-technical skills, instead of team evaluation. Clear need for Novice Raters to be trained using the NOTSS prior to implementation/evaluation of real cases

(*continued on next page*)

Table 1 (*continued*)

Article	Scale	Population Studied	Methodology	Reliability	Validity	Comments
Yule et al. [19]	NOTSS	Pre-Recorded Surgeon Behaviours – to test the NOTSS rating system	Experimental Study – with simulated operating room scenarios demonstrating a range of good to poor behaviours	Ratings between novice and expert raters for acceptable/unacceptable behaviours varied from 63 to 82% Good inter rater reliability >0.7 for only Leadership and Communication and Teamwork. Weaker for the SA and Task Management Categories.		Task Management Category Dropped from original NOTSS, leaving 4 categories. NOTSS requires minimal training as shown here (2.5hrs), however, lacks strong consistency when compared to expert ratings Interrater reliability is acceptable, and overall further training may be required prior to implementing NOTSS
OTAS						
Undre et al. [32]	OTAS Task Checklist and Behaviour rating	Operating room Teams	Observational non randomized study using non crisis urological procedures. Each subteam scored separately for the preop and post op phases.	Acceptable interobserver expert rater reliability with regards to teamwork behaviours was found for all ($r > 0.0.5$) but communication related behaviours (0.35)	N/A	Procedure specific task checklist is required + teamwork behaviour rating Focuses mainly on Teamwork related behaviours Uniquely evaluates the operative and peri-operative phases of patient care
Sevdalis et al. [31]	OTAS	Operating room Teams	Prospective Observational operating room urological cases – scored in real time	Good inter-rater reliability amongst the expert/expert pair (correlation coefficients of 0.51–0.94, for 12/15 behaviours), but not good for the expert/novice pair (0.52–0.6, for 3/15 behaviours).	Good construct validity since scores between experts vs novices demonstrate consistency.	Novices need appropriate training in order to reliably rate teamwork behaviours using the OTAS.
Undre et al. [30]	OTAS Task Checklist and Behaviour Rating scale	Operating room Teams	Observational non randomized study using non crisis general surgery procedures	N/A	N/A	The results show the OTAS rating system is feasible and useful at detecting missed steps in the pre op, operative, and post op stages.
OTTAWA GRS (Not trialed in Surgery)						
Kim et al. [33]	Ottawa Crisis Resource Management Global Rating Scale (Ottawa GRS)	First and Third Year Residents	Experimental simulation based non operating room crisis scenarios	Acceptable Interrater reliability (Intraclass coefficients of 0.59 and 0.613)	Senior residents obtained higher scores than jr. r esidents – demonstrating construct validity Review by crew resource management instructors across Canada and peer review of case scenarios – therefore appears to have content validity	Results demonstrate acceptable reliability and validity of a rating system, useful for use in the peri opeartive care setting.. Did not evaluate team or operating room performance. Used CRM and acute care physicians as raters.
Kim et al. [34]	Ottawa CRM Checklist and Ottawa GRS	First and Third Year Residents	Experimental simulation based non operating room crisis scenarios	Acceptable Interrater reliability (Intraclass coefficients of 0.59 and 0.613) for Ottawa GRS compared to the Ottawa CRM Checklist (Intraclass coefficient of 0.63 and 0.55)	Showed good correlation of Ottawa GRS with the Ottawa CRM checklist scoring	Users preferred using the Ottawa GRS due to ease of scoring, presence of an overall score and due to potential for formative evaluation Used CRM and acute care physicians as raters.

methods were employed so that surgeons could recall operative events and behaviours. The goals of the interview were to uncover cognitive skills, but also bring attention to interpersonal skills being used by surgeons during critical incidents in the operating room. Once completed, three pairs of psychologists analyzed these interviews in an independent manner using the line-by-ling coding technique from grounded theory, until an acceptable inter-rater reliability was achieved. This analysis along with a systematic program design already used in anaesthesia and in European civil aviation, helped develop the NOTSS system [17,18]. Additionally, data from a literature review of errors in the OR, attitude surveys of OR personnel, analysis of mortality reports and operating room observations, guided the development of the system.

Interviews of consultant surgeons revealed five main categories of NTS - **situation awareness, decision making, task management, leadership and communication and teamwork.** Within each of these categories there are 2–3 specific non-technical skill elements exist that expand on the overlying theme, resulting in a total of 14 separate skill elements. Furthermore, Yule et. al.worked closely with 16 consultant surgeons, to come up with examples of good and poor performance behaviours to guide raters while using the NOTSS tool [15] (Appendix 1). For instance within the category of situation awareness, one of the elements is gathering information and within this element ensuring all the relevant investigations have been reviewed is an example of an observable behavioral marker indicating good performance. Conversely arriving late in the operating room or being called to the OR repeatedly is a poor behavior. Using these observable behaviours each element is scored on a 4 point scale with 4 being good and 1 being poor, and NO indicating that the behaviour was not required for the particular case/scenario. The sum of each element rating gives the total objective score indicative of the surgeons' non-technical skills.

NOTSS testing

NOTSS reliability and sensitivity as an assessment tool was tested using 11 simulated operating room scenarios [19]. The scenarios were video recorded and employed a patient simulator, practicing surgeons, anesthetists and nurses acting in the main roles. They were designed by two surgeons, two psychologists and an anaesthetist with NTS training experience and displayed a range of non-technical skills from good to poor to assess the reliability of the NOTSS system across a range of different clinical scenarios. The majority (84%) of these participants had some experience in assessing trainee performance. Once recruited, the consultant surgeons received a brief training course (2.5 h) in human factors and in observing and rating behaviours using NOTSS.

Validity

Since the rating scale was developed using task analysis, interviews, and literature reviews, it is thought to have content validity. Furthermore, since it can differentiate between good and poor behavioral performance, it has good construct validity.

Sensitivity

A set of 'reference ratings' (determined by the scenario designers with up to 10 years experience in behavior and non-technical and technical skills rating) was used to test the sensitivity, or how closely the participants' ratings detected good and poor behaviours compared to expert ratings. There was no acceptable level of sensitivity stated by the authors, even though all participant ratings were found to be within 1 point of expert ratings. Sensitivity was shown to be highest in the task management category and lowest for situational awareness, which may be one of the most significant categories clinically. Additionally, when the rating scale was collapsed from a 4-point scale to a 2 point scale and categories were assessed as either "acceptable" or "unacceptable", a range of 63%–82% of raters agreed with expert ratings, for each of the NOTSS categories.

Reliability

Reliability was shown to be high between category and element ratings demonstrating that the different skill elements and their behavioral markers were indeed linked to their categories. Inter-rater reliability was only at acceptable levels for two categories (Leadership and Communication and Teamwork) and when using intra class correlation as an additional indicator of reliability no category displayed acceptable levels of inter-rater reliability. Since, raters were from different surgical specialties, the reliability for each specialty was also measured separately. Greater variability was found amongst scores from general surgeons than orthopedic surgeons in this study. During the final system development stages, the Task Management category was dropped, and relevant behaviours incorporated into other categories. This was due to poor inter-rater reliability for this category, and from feedback from surgeons that task management was mainly assigned during pre-operative planning rather than intra-operatively. Furthermore, it simplified the assessment tool and had overlap with elements and behaviours found within the Situation Awareness category.

Further studies by Yule et al. compared NOTSS assessment scores between novice (consultant surgeons who received only 2.5 h of NOTSS training) and expert raters (surgical team members with 10 years expertise in behaviour rating and assessment of technical and non-technical skills) [20]. The results showed that up to half of novice raters disagreed with expert ratings of video recorded simulated clinical scenarios, and generally had lower scores in most cases, when they disagreed. These discordant scores between expert and novice raters are more common in the subtler ranges of NTS performance whereas for clearly unacceptable and unsafe behavior the scores correlate more significantly. This is likely due to a lack of experience in behavioural rating, amongst the novice rater group. Additionally, the disagreements generally occurred in the mid-range behaviour categories, where the video being judged did not have an obviously outstanding behaviour or an extremely negative behavioural performance. Novice raters were uncertain how to judge the specific behaviour category when a task may have been performed poorly, but did not affect global patient care.

Use in training

The NOTSS tool was used as a basis for developing a short training course to familiarize surgeons with the non-technical skills required to enhance safety and performance in the operating theatre [21]. The session was organized and run by a team of experts involved in the NOTSS development, and consisted of three consultant surgeons, a consultant anesthetist and two industrial psychologists. A total of 13 consultants attended the event with 8 senior trainees. Eighteen surgeons evaluated the course and found it relevant to surgical practice. Most of the participants found the explicit structure and review, as well as discussion of these behaviours helpful for self-reflection and for considering how they mentor their trainees. From feedback received, a new two day course, called Safer Operative Surgery, was designed and is currently being offered by the Royal College of Surgeons of Edinburgh.

Feasibility

The NOTSS rating scale may be used to assess individuals, however only when expert raters are available. If novice raters are being used, a minimum number of hours should be spent understanding and practicing ratings prior to use in evaluation or feedback.

Non TECHnical skills (NOTECHS)

Scale development

NOTECHS was originally developed for the aviation industry for crew resource management (CRM) [12,22,23]. Its success in this field has prompted a number of adaptations initially in anaesthesia [17] and now in surgery.

NOTECHS classifies NTS into four categories, leadership & management, teamwork & cooperation, problem solving & decision making and situational awareness, with a 5 point scoring system for each category, where 1 = very poor and 5 = being very good. Empirical evidence suggests that NOTECHS can be used reliably in the context of crew resource management, prompting its trial in surgery [24,25].

Sevadalis et al. revised the NOTECHS scale so that it may be more relevant for use in the operating room [16]. They added a fifth category to the scale, called **communication and interaction.** This change comes from an understanding of differences between an operating room (OR) environment and a cockpit. The OR consists of three different professionals, namely nurses, anesthesists and surgeons, from various training backgrounds and professional cultures and therefore makes for a diverse environment. Overall this adds to decreased cohesiveness and increased technical diversity. Additionally, members of the OR team may be new to each other and often operate as ad-hoc teams due to shift rotas in situations where they haven't worked with one another before. Finally, the addition of the **communication and interaction** category was grounded in a theoretic framework of teamwork in real operating rooms developed by Healey et al. [26]. Within this framework completion of key communication related tasks and adequate performance of communication related behaviours are integral parts of an overall assessment of teamwork during real time procedures. The revised scale is rated on a 6 point scale, with 1 = not done to 6 = done very well.

Validity

Since the scale was adapted from extensive work demonstrating it's validity in aviation, the NOTECHS is thought to have content validity as a human factors measurement tool. In surgery, it has been shown to differentiate between good and poor behaviors and as such has good construct validity. [16]

Reliability

The revised NOTECHS was tested using simulation based operating room crisis scenarios. [16] The scale was found to have good reliability, with Crohnbach Alpha > 0.7 for all 5 subscales of the revised NOTECHS. Additionally, ratings between trainers and trainees were similar, suggesting that NOTECHS may be used as an effective self assessment tool. Reliability was also good with repeated administrations of the scale before and after non-technical skills training. Finally, the scale was reliable in the face of consecutive administrations and equally effective when used by nurses, anesthetists and surgeons.

Use in training

NOTECHS use has progressed into the clinical setting. A version adapted to measure teamwork performance in the OR, the Oxford NOTECHS system, with the original four categories has demonstrated some intriguing results. In a study by Mishra et al. [27], non-technical and technical skills were compared while surgical teams performed 26 laparoscopic cholecystectomies and 22 carotid endarterectomy procedures. The observation clinical human reliability assessment (OCHRA) tool was used to measure technical errors. Surgeons, nurses and anesthetists were observed and scored independently while operating by two observers, a surgical trainee and a human factors expert. The mean scores for surgeons, anesthetists, and nurses out of a maximum of 16 were 13.3, 11.4 and 10.8, respectively. Ratings between expert and novice raters showed good reliability as per Crohnbach alpha of 0.88. Errors outside of the operative field, for instance malfunction of essential equipment or inability to set up an instrument, were recorded as Non-Operative Procedural Errors (NOPEs). A safety attitudes survey, operation time, length of patient hospital stay and other outcomes were also recorded. The participants received non-technical skills training between two separate operative procedures.

The findings suggest that non-technical skill scores remained unchanged for surgeons even after human factors skills training, likely due to a high level performance at baseline. The technical performance and the number of NOPE's decreased after non-technical skills training.

Though the overall correlation between technical errors and non-technical skills was weak, there was a strong correlation between a surgeon's situational awareness and technical error rates. The operating times and patient length of stay (LOS) after the operation remained unchanged after non-technical skills training.

Overall, these findings suggest that human factors training improves non-technical skills and decreases errors in the OR, likely due to improvements in teamwork and communication. However, due to a small sample size in this study, these improvements did not translate in to tangible decreases in patient hospital LOS or operating times, and future large scale studies will be needed to evaluate these outcomes.

Feasibility

The NOTECHS rating scale evaluates a wide array of non-technical skills. It provides good behavioural descriptors for each of the categories and as a result requires less training prior to use. Furthermore, since this scale demonstrates reliability between trainee and expert raters it can be used for self assessment at a significantly lower cost, compared to other ratings scales. Additionally, it can be used for team and individual evaluation/feedback and be used in real time, as has been demonstrated in preliminary studies.

OTAS (Observational Teamwork Assessment for Surgery)

Scale development

Developed in 2006, the OTAS came about from a need to have a wider assessment of the factors that play a role in patient outcomes [30]. Additionally, an emphasis was placed on evaluating teamwork in the OR that went beyond individual teamwork skills. Using a basic input-output model of team performance that has been established in aviation, and is prominent in team theory literature, a model for surgical teamwork was created.

OTAS has two elements of evaluation, the first is a task checklist and the other is a team behavioural assessment. The task checklist was constructed using practice guidelines, theatre protocols and expert advice. Tasks are categorized into either patient, equipment or communication related actions, and are scored as a 'yes' or 'no' according to team performance. The OTAS evaluation is observational, and a task is considered to be complete only if the task is performed explicitly.

Behaviours were assessed on a set of teamwork behaviours that included shared monitoring, communication, cooperation, co-ordination and shared leadership, adapted from Dickinson and McIntyre's model of teamwork. Behavioural dimensions were also guided by interviews done by Undre et al. and from other measures of teamwork, such as those used by Fletcher et al. [8] to rate anesthetists non-technical skills. Team performance was scored using a 7-point Likert scale.

Measures of good team performance were derived from guidelines of best surgical practice and combined with broader dimensions of behavior to assess specific tasks in the operating room. The team

performance checklist focuses on routine operating room scenarios within General Surgery and Urology, and encompasses all aspects of surgery, including the pre and post operative phases of surgery.

Validity

The OTAS task checklist and behavior rating scale have demonstrated good construct validity since expert and novice raters for each demonstrate consistency amongst themselves. However, further testing with respect to content validity needs to be done.

Reliability

Sevdalis et al. have studied the use of the OTAS scale by both expert and novice raters to rate teamwork in the operating room. The results demonstrate a high level of correlation amongst expert raters, while a poor correlation between novice and expert raters. The findings demonstrate that novice raters must be trained and their learning curve determined prior to implementing the OTAS scale in trainee assessment.

Use in training

Undre et al. have used the OTAS tool to assess teamwork performance for urological and general surgical procedures, including oncology cases such as colectomies, anterior resections, gastrectomy, and Hartmann's procedure[31, 33]. The study showed that all members of the OR team had low scores on communication behaviours. Furthermore, surgeons' behavioural scores deteriorated towards the end of procedures. The latter may occur if the surgeon leaves the OR towards the end of the procedure, and designates a junior surgeon to complete the remaining tasks, such as skin closure. Additionally, failure to check equipment or confirm the procedure verbally occurred frequently, while delays and changes to procedures occurred in roughly two-thirds of all cases. Overall, the OTAS helps identify areas of behavioural deficiencies in an OR setting, and also identifies areas where equipment/task failures occur, such that they can be identified and corrected.

Feasibility

The use of OTAS system as a rating tool is limited since the task specific checklist is limited to only certain procedures. Moreover, since it relies on an active real time observer to rate both behaviours and tasks it may not be a cost effective or feasible rating methodology in real operating room situations. The OTAS is a more comprehensive rating tool and may be beneficial to assess system wide actions and team interactions, instead of being limited to the operating room setting alone.

Other rating scales

Other evaluation tools developed in the context of Anaesthesia and critical care provide valuable lessons, and will help formulate future non-technical training models in surgery. One such tool is the Ottawa GRS, which uses five categories of CRM skills and was developed using a Delphi process [33,34]. It received input from critical care physicians and acute care specialty physicians from emergency medicine and anaesthesia. The five categories are **problem solving, situational awareness, leadership, resource utilization, and communication**. A score of 1–7 is assigned for each category, and descriptors are provided to guide scoring within each of these category. The amount of cues necessary for residents to act is also taken into account in the Ottawa GRS.

A recent study employing simulated crisis scenarios was used to compare the Ottawa GRS to a traditional checklist type evaluation system. Residents with prior simulator experience in residency were excluded. 59 residents were recruited, 21 from internal medicine, 15 from surgical specialties, 16 from family medicine and 3 from anaesthesia and emergency medicine. These residents were either in their first year or third year of their training. They were expected to manage simulated scenarios, such as a cardiac event in a post operative patient or acute shock and respiratory failure after a severe trauma. Sessions were rated by experts 2 days or 2–3 weeks after a simulator tutorial session. The findings suggest that the Ottawa GRS is a valid rating scale and can differentiate the CRM performance amongst residents. Furthermore, inter-rater reliability was high when the Ottawa GRS was used in these patient crisis care scenarios. Raters using the scale acknowledged the flexibility, ease of use and simplicity of the Ottawa GRS. Overall, the tool was found to be valid and feasible for non-technical skills of surgical residents.

Evaluation tools in anaesthesia have been developed from CRM work in aviation, and have resulted in training courses for anaesthesia trainees. The Anaesthetists Non-Technical Skills (ANTS) [8] scale, was developed from expert surveys, study of adverse events in anaesthesia and consultant feedback. The tool has four broad categories subdivided further into 15 elements. The categories are **task management, team working, situation awareness and decision making** and behavioural markers to illustrate good/poor behaviours are present in each category. Its application has required minimal training for raters and has acceptable levels of validity, reliability and usability. The ANTS system has allowed the development of various non-technical skills training courses.

Surgical oncology presents additional challenges in terms of NTS to most other surgical services, whereas the clinical decisions are made almost exclusively by senior surgeons in most surgical patients, cancer patient care decisions are increasingly made by multidisciplinary teams (MDT) who consist of diverse healthcare professionals from varying backgrounds and different perspectives on ideal patient care. Although this method of patient management is the preferred modality in many countries including the UK, there remains little research on its impact on patient outcomes [35].

The MDT model adds new dimensions to decision making and leadership as decision making shifts from primarily individual decisions to decisions reached by consensus and where leadership is often unclear. Lamb et al. discussed the non-technical factors influencing MDT effectiveness, finding that as well as teamwork and leadership other factors such as team professional diversity can positively influence outcomes and that while theatre time is obviously protected, in many organizations MDT time is not and attendance due to scheduling conflicts can be a key component to determining effectiveness [35].

Furthermore more while it has been demonstrated that while clear leadership is required for a team to function properly [36], perhaps counterintuitavely it has been demonstrated that good leadership in MDTs is associated with a number of leaders and is impaired by having just one individual as leader [37].

Consensus decision making in the non hierarchical MDTs is one of its purported advantages, however the issue of disagreements in this setting is further complicated by the finding that members are often unwilling to dissent even when they have strong disagreements with the decisions being made [38], this presents a requirement for a robust system to formally check for disagreements and address these accordingly as the majority of the team will be unaware of any disagreements let alone how to address them.

Discussion

As we move forward, surgical education will continue to evolve as a result of reductions in work hours, increasing requirements for organizational efficiency, and an increased focus on patient safety. As a result of these pressures, surgical curricula need to design and implement innovative models to evaluate, train and develop incoming residents. In surgical oncology, the role of the surgeon in

communicating and planning patient care plans is growing. However, previous studies suggest that the current training curriculum does not reflect this and that better educational models to teach non-technical skills are needed [39,40]. For the surgical resident, skills not often taught in medical school, become essential in coping and effectively managing the many demands of the operating room, the surgical ward and the emergency care setting.

The present paper reviews the literature on non-technical skills as it relates to surgery. The two major evaluation tools restricted to measuring behaviours in the operating room are the NOTSS and the NOTECHS systems. An additional rating scale called OTAS that assesses team performance behaviours in the OR and combines this with a procedural task checklist, was also reviewed. The NOTSS and NOTECHS scales claim validity, however, there has been little empirical work to study the validity of NOTSS or NOTECHS in the surgical realm, i.e. do these two scales capture what they purport to capture. Therefore, moving forward this is an area of focus that needs further investigation. The OTAS scale evaluates a limited number of teamwork behavioural categories compared to the NOTSS and NOTECHS, but does offer the addition of a peri-operative task checklist that may have applicability in more complex surgical cases, such as in oncology.

In terms of reliability, the revised NOTECHS system is a reliable evaluation tool based on current studies, suggesting consistency when used by evaluators from different educational backgrounds, under varying circumstances (pre and post training). The revised NOTECHS has not been tested for sensitivity (i.e. how close the rater scores were to expert scores), and therefore, this can be an avenue of future studies. The OTAS scale should be limited to use by expert raters, as it exhibits poor correlation amongst the novice and expert rater pairs. The NOTSS scale underwent sensitivity and reliability testing with conflicting results; hence, further testing is required prior to its use in surgical evaluation and feedback.

In surgical oncology, peri-operative planning is vital and requires a team based approach to patient care which presents its own challenges in terms of NTS in addition to intra-operative requirements. The OSATS study by Undre et al, pointed to frequent failures while both confirming procedures verbally and with surgical/anaesthetic equipment. Additionally, delays and changes occurred in a majority of cases, potentially compromising patient care. These could have an adverse impact on patient outcomes and should be a focus of future studies focusing on oncologic procedures. Furthermore, new laparoscopic techniques and innovative technologies in surgical oncology present a learning curve for surgeons and are therefore associated with a higher error rate [41]. Ensuring appropriate teamwork, communication, situational awareness and decision making in these situations can help buffer the effects of the early adoption phase and lead to better patient outcomes and improved learning.

In other high reliability organizations, raters receive extensive training on all relevant aspects of non-technical skills and are taught to distinguish good and poor behaviours before ever assessing actual performance. The training received by raters in the NOTSS study however was only 2.5 h, which may have resulted in inaccurate assessment and consequently poor sensitivity and reliability scores. Second, since simulation sessions were only presented for a brief period of time, it may not have given ample opportunities for all behaviours to be rated adequately. Future work should assess the NOTSS scale in real time simulation scenarios using operating teams, consisting of nurses, anesthetists and surgeons. It is notable, that the NOTECHS tool on the other hand, does not require significant training prior to being used by evaluators, thus making it a practical assessment tool for surgical educators.

Surgical residents are certainly required to use non-technical skills in the operating room environment, but these skills are also essential outside it. On the wards, residents often face crisis scenarios, involving post operative complications, transfusion reactions, or other medical management challenges. How they react to these situations in the face of demands in the OR, consults in the ER and with various multidisciplinary team members has not been studied. Previous experience, along with skills such as collaboration, professionalism, leadership and communication are vital to decision making and ultimate patient management. These skills would be difficult to evaluate in real time, without posing a risk to patients. Therefore, future endeavors must explore the use of surgical simulators and multidisciplinary teams to train and evaluate these non-technical skills to surgical residents. Residents across the board must strive to develop these skills, just as they are encouraged to master the surgical literature and become technically proficient.

Evaluation tools must ultimately help residents receive feedback and training with respect to their non-technical skills. A CRM type classroom and simulator session training course has been formulated in anaesthesia, called the Crisis Avoidance and Resource Management for Anaesthetists (CARMA) [35]. Theory is delivered through formal presentations and themes were explored through the use of case based discussions and small group exercises. The feedback of anesthetists and psychologists was used in course development. Overall the course has received very positive evaluations, and areas highlighted for improvement were communication, team working and reviewing cases aloud.

Surgical trainees might benefit from similar simulated operative, and post operative scenario training sessions. By placing trainees in environments that are similar to those studied in CARMA courses, residents will be evaluated, receive feedback and improve their non-technical skills.

Just as surgical skills labs have been constructed to practice knot tying, suturing and laparoscopic skills, training facilities to develop non-technical skills must be envisioned. These would consist of a psychologist, a surgeon and a multidisciplinary team that allows surgeons to train and develop skills essential to leading and managing a team effectively. These constructs would offer a series of training modules that focus on each specific element of non-technical skills in the context of a simulated patient crisis scenario. Ultimately, an evaluation tool would have a knowledge base category alongside non-technical skills, and residents at each year of residency training would be able to monitor their progress through their years of training. This would change the caliber of future surgeons and ensure they are well-rounded communicators, decision makers, and leaders along with being skilled operators, resulting in improved patient outcomes.

Acknowledgement

Rajesh Aggarwal is funded by a Clinician Scientist Award from the National Institute of Health Research, Department of Health, UK.

Appendix. NOTSS Evaluation Form

Postop Complication Scenario Evaluation Sheets	Ellement Rating (1–4)	Feedback on performance and debriefing notes
Communication and Teamwork		
Exchanging information		
Establishing a shared understanding		
Co-ordinating team activities		
Decision Making		

(*continued*)

Postop Complication Scenario Evaluation Sheets	Ellement Rating (1–4)	Feedback on performance and debriefing notes
Considering options		
Selecting and communicating option		
Implementing and reviewing decisions		
Situation Awareness		
Gathering information		
Understanding information		
Projecting and anticipating future state		
Leadership		
Setting and maintaining standards		
Supporting others		
Coping with pressure		

References

[1] Calland JF, Guerlain S, Adams RB, Tribble CG, Foley E, Chekan EG. A systems approach to surgical safety. Surg Endosc 2002;16:1005–14.

[2] Healey AN, Sevdalis N, Vincent C. Measuring intraoperative interference from distraction and interruption observed in the operating theater. Ergonomics 2006;49:589–604.

[3] Lingard L, Garwood S, Poenaru D. Tensions influencing operating room team function: does institutional context make a difference. Med Educ 2004;38: 691–9.

[4] Lingard L, Reznick R, Espin S, Regehr G, DeVito I. Team communications in the operating room: talk patterns, sites of tension, and implications for novices. Acad Med 2002;77:232–7.

[5] 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;13:330–4.

[6] Helmreich RL, Merritt AC, Wilhelm JA. The evolution of Crew Resource Management training in commercial aviation. Int J Aviat Psychol 1999;9:19–32.

[7] Flin R, Martin L. Behavioural marker systems in aviation. Int J Aviation Psychol 2001;11:95–118.

[8] 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(5):580–8.

[9] Bogner M, editor. Human error in medicine. Hillsdale, NJ: LEA; 1994.

[10] Bogner M, editor. Misadventures in Health care. Mahwah, NJ: LEA; 2004.

[11] Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. Surgery 2003;133:614–21.

[12] Klampfer R, Flin R, Helmreich RL, Hausler R, Sexton B, Fletcher G, et al. Enhancing performance in high risk environments: recommendations for the use of behavioral markers. Gottlieb Daimler and Karl Benz Foundation Kolleg Group Interaction in High Risk Environments (GIHRE), http://www.siaa.asn.au/simtect/2005/GIHRE2.pdf; 2001 (assessed on 16.04 08).

[13] Baker D, Mulqueen C, Dismukes R. Training raters to assess resource management skills. In: Salas E, Bowers C, Edens E, editors. Improving teamwork in organizations. Mahwah, NJ: Lawrence Earlbaum; 2001. p. 131–45.

[14] Flin R, Yule S, Paterson-Brown S, Maran N, Rowley D, Youngson G. Surgeons' non technical skills. Surg News 2005;4:83–5.

[15] Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D. Development of a rating system for surgeons' non-technical skills. Med Ed 2006;40:1098–104.

[16] Sevdalis N, Davis RE, Koutantji M, Undre S, Darzi A, Vincent CA. Reliability of a revised NOTECHS scale for use in surgical teams. Am J Surg 2008;196:184–90.

[17] Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Rating non-technical skills: developing a behavioural marker system for use in anaesthesia. Cogn Technol Work 2004;6:165–71.

[18] O'Connor P, Flin R, Fletcher G. Methods used to evaluate the effectiveness of flight crew CRM methods in the UK aviation industry. Hum Factors Aerospace Saf 2002;2:235–55.

[19] Yule S, Flin R, Maran N, Rowley DR, Youngson GG, Paterson-Brown S. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behaviour rating system. World J Surg 2008;32:548–56.

[20] Yule S, Rowley DR, Flin R, Maran N, Youngson G, Duncan J, Paterson-Brown S. Experience matters: comparing novice and expert ratings of non-technical skills using the NOTSS system. ANZ J Surg 2009;79(3):154–60.

[21] Notss course for training surgeons. 86–89. Surgeon; April 1, 2007.

[22] Flin R, Martin L, Goeters KM, Hörmann H-J, Amalberti R, Valot C, et al. Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. Hum Factors Aerospace Safe 2003;3:97–119.

[23] O'Connor P, Hormann HJ, Flin R, Lodge M, Goeters KM. Developing a method for evaluating Crew Resource Management skills: a European perspective. Int J Aviat Psychol 2002;12:263–85.

[24] Moorthy K, Munz Y, Adams S, Pandey V, Darzi A. A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. Ann Surg 2005;242:631–9.

[25] Moorthy K, Munz Y, Forrest D, Pandey V, Undre S, Vincent C, et al. Surgical crisis management skills training and assessment. Ann Surg 2006;244:139–47.

[26] Healey AN, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. Qual Safe Health Care 2004;13(suppl. 1):i33–40.

[27] 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.

[28] 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 threatre. Qual.Saf Health Care 2009;18:109–15.

[29] Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating threatre. Qual Saf Health Care 2009;18:104–8.

[30] Undre S, Healey AN, Darzi A, Vincent CA. Observational assessment of surgical teamwork: a feasibility study. World J Surg 2006;30:1774–83.

[31] Sevdalis N, Lyons M, Healey A, Undre S, Darzi A, Vincent C. Observational teamwork assessment for surgery construct Validation with expert versus novice raters. Ann Surg; 249: 1047–1051.

[32] Undre S, Sevdalis N, Healey A, Darzi A, Vincent A. Observational teamwork assessment for surgery (OTAS): Refinement and application in urological surgery. World J Surg 2007;31:1373–81.

[33] Kim J, Neilipovitz D, Cardinal P, Chiu M, Clinch J. A pilot study using high-fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients: the University of Ottawa Critical Care Medicine, high-fidelity simulation, and crisis resource management I study. Crit Care Med 2006;34:2167–74.

[34] Kim J, Neilipovitz D, Cardinal P, Chiu M. A comparison of global rating scale and checklist scores in the Validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated Emergencies. Simulation in Health Care 2009;4:6–16.

[35] Lamb B, Green JSA, Sevdalis N. Decision-making in surgical oncology. Surgical Oncology 2011;20(3):163–8.

[36] Borrill C, West M, Shapiro D, Rees A. Team working and effectiveness in healthcare. Br J Healthc Manage 2000;6(8):364–71.

[37] Haward R, Amir Z, Borrill C, Dawson J, Scully J, West M, et al. Breast cancer teams: the impact of constitution, new cancer workload and methods of operation on their effectiveness. Br J Cancer 2003 Jul 7;89(1):15–22.

[38] Sidhom MAP, Poulsen M. Group decisions in oncology: Doctors' perceptions of the legal responsibilities arising from multidisciplinary meetings. J Med Imaging Radiat Oncol 2008;52(3):287–92.

[39] Sariego J, Byrd M, Matsumoto T, Kerstein M. The changing role of the surgeon in the management of cancer in the University setting and its impact on resident training. J Surg Oncol 1992;51:143–6.

[40] Hoffman M, Ferri J, Sison C, Roter D, Schapira L, Baile W. Teaching communication skills: an AACE survey of oncology training programs. J Cancer Educ 2004;19(4):220–4.

[41] Al-Akash M, Boyle E, Tanner WA. Training on N.O.T.E.S.: from history we learn. Surg Oncol 2009;18(2):111–9.