



# The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre

P McCulloch,<sup>1</sup> A Mishra,<sup>1</sup> A Handa,<sup>1</sup> T Dale,<sup>2</sup> G Hirst,<sup>2</sup> K Catchpole<sup>1</sup>

## See Commentary, p 91

► Additional material is published online only at <http://qshc.bmj.com/content/vol18/issue2>

<sup>1</sup>Nuffield Department of Surgery, University of Oxford, Oxford, UK; <sup>2</sup>Atrainability Ltd, Guildford, UK

Correspondence to: Mr P McCulloch, Nuffield Department of Surgery, 6th Floor, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK; [peter.mcculloch@nds.ox.ac.uk](mailto:peter.mcculloch@nds.ox.ac.uk)

Accepted 30 December 2008

## ABSTRACT

Unintended harm to patients in operating theatres is common. Correlations have been demonstrated between teamwork skills and error rates in theatres. This was a single-institution uncontrolled before–after study of the effects of “non-technical” skills training on attitudes, teamwork, technical performance and clinical outcome in laparoscopic cholecystectomy (LC) and carotid endarterectomy (CEA) operations. The setting was the theatre suite of a UK teaching hospital. Attitudes were measured using the Safety Attitudes Questionnaire (SAQ). Teamwork was scored using the Oxford Non-Technical Skills (NOTECHS) method. Operative technical errors (OTEs), non-operative procedural errors (NOPEs), complications, operating time and length of hospital stay (LOS) were recorded. A 9 h classroom non-technical skills course based on aviation “Crew Resource Management” (CRM) was offered to all staff, followed by 3 months of twice-weekly coaching from CRM experts. Forty-eight procedures (26 LC and 22 CEA) were studied before intervention, and 55 (32 and 23) afterwards. Non-technical skills and attitudes improved after training (NOTECHS increase 37.0 to 38.7,  $t = -2.35$ ,  $p = 0.021$ , SAQ teamwork climate increase 64.1 to 69.2,  $t = -2.95$ ,  $p = 0.007$ ). OTEs declined from 1.73 to 0.98 ( $u = 1071$ ,  $p = 0.009$ ), and NOPEs from 8.48 to 5.16 per operation ( $t = 4.383$ ,  $p < 0.001$ ). These effects were stronger in the LC group than in CEA procedures. The operating time was unchanged, and a non-significant reduction in LOS was observed. Non-technical skills training improved technical performance in theatre, but the effects varied between teams. Considerable cultural resistance to adoption was encountered, particularly among medical staff. Debriefing and challenging authority seemed more difficult to introduce than other parts of the training. Further studies are needed to define the optimal training package, explain variable responses and confirm clinical benefit.

We studied personnel performing laparoscopic cholecystectomy (LC) and carotid endarterectomy (CEA) procedures in the theatre suites of a major UK teaching hospital. These procedures were selected to minimise difficulties in evaluating performance caused by variations in the operative procedure. All CEA operations were performed in one theatre while the LC procedures were mostly done in three others, one of which was designated a day-case theatre. All CEA operations were elective or semielective procedures performed or directly supervised by consultant vascular surgeons. About 20% of LC were performed by experienced trainees without supervision, and a similar percentage were “hot” procedures. There

were no major differences in availability and quality of equipment between the LC theatres except that intraoperative cholangiography was difficult to perform in the day-case theatre. The consultant surgical and anaesthetic staff and senior theatre nurses were largely constant throughout the study, but junior staff turnover was relatively high, typical of UK practice. The day-case theatre was situated in a different hospital from the other theatres; the nursing staff in this theatre were predominantly trained in the Philippines, whereas most staff in the other theatres were UK-trained. The core CEA team had considerable experience of working together, and the senior members had all been trained by the same senior vascular surgeon when the CEA operation was first practised regularly at the hospital. We aimed to study consecutive cases but made exceptions due to refusal of consent (one case) or unavailability of research staff (three cases). After the intervention, we also excluded cases where there were fewer than two members of the theatre team who had attended the classroom part of the training course (six cases).

## PROBLEM

Current evidence suggests that modern healthcare causes unintentional harm to between 3% and 16% of hospital inpatients.<sup>1,2</sup> The factors implicated include organisational complexity, reliance on potentially dangerous high-technology equipment and the lack of systematic communications and teamwork training for staff.<sup>3,4</sup> The operating theatre is reportedly the environment where patient harm is most likely.<sup>5</sup> Theatre work is perceived as stressful by staff,<sup>6</sup> and communication between team members as flawed and difficult.<sup>7</sup> Reports on high-profile disasters in surgery such as wrong organ removal or the Bristol cardiac surgery incident have stressed the major role of teamwork and communication problems in bringing these about.<sup>8</sup> Reviews of practice in high-reliability industries such as aviation, nuclear power and offshore oil production emphasise the need for training in principles of teamwork and free communication to reduce the risks of error,<sup>9</sup> and studies in healthcare show that such training improves safety culture and attitudes across a range of specialty settings.<sup>10,11</sup> We were aware that, like most large hospitals, our local Trust had experienced major intraoperative incidents such as wrong organ removal in the recent past. Observational studies in paediatric cardiac surgery have shown (a) a clear association between the

accumulation of minor errors in an operation and the occurrence of a major incident with potentially serious implications for the patient<sup>12</sup> and (b) a convincing correlation between the quality of theatre team “non-technical skills” and the number of technical errors which occurred.<sup>13</sup> The environment in which these studies were conducted is highly unusual, and in particular combines extremely high risks of adverse or fatal outcome with an absolute need for close cooperation between anaesthetists, surgeons and perfusionists in the heart–lung bypass process. A study of non-technical skills and technical errors in LC in our theatre suite, however, showed evidence of deficiencies in both technical and non-technical skills of the same order as that seen in the cardiac environment, and confirmed the association between them.<sup>14</sup> The hypothesis that improvements in theatre team non-tech skills could substantially reduce errors and adverse outcomes of surgery has become widely accepted, but the evidence supporting this hypothesis is currently scanty.

Studies of theatre teamwork, communication and safety have been largely observational, focusing on the development of methodology<sup>15–17</sup> and on the correlations between non-technical skills and adverse events.<sup>13 18 19</sup> Only one major interventional study has attempted to evaluate non-technical skills training in hospitals,<sup>20</sup> reporting a reduction in clinical errors in emergency rooms. No similar study of theatre teams has been reported. We felt that the postulated benefits of non-tech skills training for theatre teams merited further study. Based on our reading, we expected we might, through aviation-style “CRM” teamwork training, be able to improve teamwork in theatre teams, and thereby reduce technical errors and misunderstandings. We therefore hoped to see reductions in the number of potentially significant errors and mishaps in theatre with potential for harm to patients, and perhaps to observe improvements in some clinical outcome measures. We therefore undertook a single-institution before–after study of a non-technical skills team training intervention.

### KEY MEASURES FOR IMPROVEMENT

We first observed performance and outcomes in a series of operations, then introduced a training intervention, and finally conducted a second observation period. The three phases lasted 6, 3 and 6 months respectively. We initially agreed to observe a theatre team from another specialty without providing an

#### Box 1 Example (for surgical subteam situation awareness) of the anchoring scoring criteria for one Non-Technical Skills Scale dimension

- 1: Below standard. Surgeon fixated on operative field, completely unaware of problems “at the anaesthetic end” (eg, in marked bradycardia, has to be told repeatedly to stop).
- 2: Basic standard. Shows a level of awareness adequate if things are running smoothly, but could be detrimental (eg, foot left on diathermy pedal). Assistant reveals poor understanding of patient’s anatomy.
- 3: Standard. Periodically gathers information about surroundings and activities of other team members.
- 4: Above standard. Aware of patient condition at all times, and is able to predict potential problems. Appreciates stage of operation and ensures equipment is ready before it is needed. Aware of who is present in theatre at all times, and shows anticipation in changes

intervention, but this team withdrew after some initial observations due to sensitivities over their perception that involvement called into question their current performance. The study was therefore an uncontrolled before–after intervention study. Our hypothesis postulated an effect of teamwork training starting with improved attitudes, knowledge and performance in terms of teamwork and communication, leading to a reduction in “unsalvaged” technical errors, misunderstandings and omissions, and via this route to less imperfect technical results and ultimately less complications. We therefore designed our outcome measures to evaluate a “chain of evidence” linking teamwork training via improved teamwork attitudes, skills and performance to reduced technical errors, untoward events and adverse outcomes. We therefore measured attitudes, knowledge and behaviour relating to workplace teamwork and safety, and recorded technical and procedural errors by the theatre team members. We also counted complications and critical incident reports and measured operating time and length of hospital stay. Ethical approval for the study was obtained from the Milton Keynes LREC (ref no 04/Q1603/35). Informed consent was sought from all staff involved in the operations studied, and from the patients undergoing them.

### METHODS FOR EVALUATION

Observations were made by researchers (AM and KC) present in theatre throughout each operation. Both had more than 2 years’ experience in theatres as a scientific observer (KC) or trainee surgeon (AM). KC had extensive human factors training and previous experience of methods for observing technical and non-technical skills: he received prior technical and anatomical training about the operations studied, including informal observation of up to 20 procedures before beginning this study. AM received intensive training in scoring non-technical skills from an aviation expert in Crew Resource Management with previous experience of research in operating theatres (TD). Observations were made as ethnographic free-form, real-time notes and written up immediately after the operation.

A single observer was able to observe non-technical skills, technical errors and procedural errors. To assess reliability, both observers recorded scores independently in a proportion of operations. To avoid observer bias in scoring teams after training, an independent third observer (PS), with no previous involvement in the project, co-observed non-technical skills in a sample of post-training cases (n = 11).<sup>21</sup>

### Non-technical skills

Non-technical skills were observed using the Oxford NOTECHS scoring system, developed from a tool designed for aviation personnel,<sup>22</sup> via previous studies of non-technical skills and surgical error.<sup>19</sup> NOTECHS classifies non-technical skills into four dimensions: (1) leadership and management (L&M), (2) teamwork and cooperation (T&C), (3) problem-solving and decision-making (P&D) and (4) situation awareness (SA). A score between 1 and 4 is assigned for each dimension by semiobjective assessment using prespecified behavioural markers, anchored to four categories (below standard; basic standard; standard; excellent (see box 1)).<sup>21</sup> Each of the theatre subteams (surgeons, anaesthetists and nurses) was scored on each dimension. Subteam performance was taken as the sum of the dimension performances (out of 16). Team performance was calculated from the sum of the subteam performance scores (out of 48). Whole team NOTECHS dimension performance was scored as the sum of all subteam performances in that

**Box 2 Case study 1: distraction**

The operation was a laparoscopic cholecystectomy performed by two senior trainees. The anatomy was complex, and the decision was made to convert to an open procedure. At this point, the senior consultant surgeon entered theatre and announced he would scrub in too. He produced his music system and proceeded to play music loudly, which made conversation between the operating surgeons and the scrub nurse difficult, and then he began a loud and detailed discussion with the anaesthetist about a forthcoming case. After a few minutes, the consultant expressed himself content with the way the registrars were progressing and left the theatre leaving the music still on. The circulating nurse then turned the volume down, and subsequently wrote "would Mr XXX please not turn the music up loud in the theatre while people are trying to work" in a comments book that had been placed in each theatre for the study. Some time later, the consultant read the comment but made light of it.

dimension (out of 12). Thus, a score was obtained in each dimension for the theatre team as a whole, and for each subteam. In 36 cases (24 LC and 12 CEA), where two observers carried out independent parallel observations, NOTECHS proved extremely reliable with  $R_{wg} = 0.983$  overall (see table 1).

**Technical errors**

Operative technical performance was assessed using the Observation Clinical Human Reliability Assessment (OCHRA) tool, developed for analysis of technical errors during laparoscopic cholecystectomy.<sup>23, 24</sup> Nine stages are recognised, each consisting of a series of subtasks. Errors were recorded as Operative Technical Errors (OTEs). We adapted this method for use in carotid endarterectomy, which was divided into nine mandatory stages and one optional stage.

**Procedural errors**

We recorded slips, errors and mishaps occurring in theatre but outside the operative field, using a taxonomy developed in earlier studies,<sup>19</sup> which we described as Non-Operative Procedural Errors (NOPEs). This category included absence or malfunction of essential equipment, inability to use or set up equipment, dropping items and miscommunications (table 2). Like the Oxford non-technical skills scoring method (NOTECHS), operative technical errors and non-operative procedural errors were also reliably observed ( $R_{wg} = 0.987$  and 0.831 respectively).

**Safety climate assessment**

All staff under study were asked to complete the Safety Attitudes Questionnaire<sup>25</sup> prior to training (after the initial 6-month observation period) and again after the completion of the coaching period (3 months after the classroom training).

The SAQ evaluates knowledge and attitudes relating to workplace safety on six dimensions, of which only one (Teamwork Climate) was directly relevant to the training delivered.

**Additional outcome data**

The operating time, from patient arrival in theatre to disconnection from anaesthetic support, was recorded for all cases. Postoperatively we recorded: length of hospital stay; any return to theatre or unplanned admission to the intensive treatment unit; critical incident reports; complications recorded in case notes or recalled by the patient during a structured telephone interview conducted a mean of 12 weeks after surgery.

**ANALYSIS AND INTERPRETATION**

We studied 26 LC operations and 22 CEA procedures before the intervention, and 32 and 23 procedures afterwards. All personnel ( $n = 83$ ) regularly involved in LC or CEA procedures were invited to the classroom sessions, and 54 (65%) attended. All personnel were involved in subsequent "bedding in" coaching in theatre. After teamwork training, at least two team members had to have attended classroom training to make the case eligible for study.

We compared SAQ scores using paired t tests; non-technical skills scores (Oxford NOTECHS), non-operative procedural errors (NOPEs) and operative duration using unpaired t tests; and operative technical errors and length of stay using Mann-Whitney U tests. Length of stay was also analysed by examining the difference in population percentages with z scores.<sup>26</sup> We evaluated agreement between observers using  $R_{wg}$ , and the relationships between technical errors and non-technical skills scores (NOTECHS) using the Spearman correlation. We also compared post-training non-technical skills scores in the presence or absence of a preoperative briefing.

**STRATEGY FOR CHANGE**

Our classroom training intervention was based on the principles of civil aviation Crew Resource Management (CRM) training. Attendees received 9 h of mixed didactic and interactive teaching. The aims of the training programme were to increase knowledge, change attitudes and improve behaviour in relation to: (a) safety, situation awareness, and error management; (b) self-awareness, communication and assertiveness; and (c) decision-making, briefing and debriefing. The three 3 h modules each concentrated on one of these aspects. In the first module, the theory of organisational error and the value of flat hierarchy, clear impersonal communication and an understanding of situational awareness and "red flags" were explained and demonstrated. In the second module, Myers-Briggs profiles were used to explain personal communication styles, and a tool for graded authority challenge was introduced. In the third module, role-play was used extensively to develop briefing and debriefing skills. A video, role-play scenarios, wall posters and

**Table 1** Interobserver reliability for NOTECHS ( $n = 36$ )

$R_{wg}$	Leadership and management	Teamwork and cooperation	Problem solving and decision-making	Situation awareness	Total
Surgeons	0.889	0.889	0.911	0.855	0.959
Anaesthetists	0.944	0.911	0.933	0.933	0.969
Nurses	0.867	0.889	0.867	0.900	0.957
Team	0.937	0.932	0.962	0.956	0.985

NOTECH, Non-Technical Skills scale.

**Table 2** Taxonomy of non-operating procedural errors

Classification of problem	Example
Absence	Circulating nurse out of theatre when she is needed to get clipper for cystic duct/artery
Coordination/communication problem	Surgeon asks nurse $\times 3$ for vascular sling before receiving it
Distraction	Mobile phone rings loudly during case
Equipment/workspace management problem	Diathermy unplugged when required
Equipment operation problem	Transducer not zeroed giving false readings
Equipment problem	Sutures break
Expertise/skill problem	Consultant surgeon captures error made by trainee surgeon
External resource problem	Piece of equipment is missing from standard set
Patient-sourced procedural difficulties	Difficult anatomy causing operative difficulties
Planning problem	Difficult intubation anticipated but not planned for
Procedure-related error	Arterial clamp time not recorded
Non-operative psychomotor error	Retractor is dropped
Resource management problem	Surgeon leaves assistant to close without confirming ability to do so
Safety consciousness problem	Anaesthetist not wearing face mask in carotid endarterectomy, despite being protocol during vascular cases
Vigilance/awareness problem	Anaesthetist fails to note significant drop in arterial pressure

pocket cards containing summaries of the course content were created and distributed (see web material). Following the classroom course, teams were supported in theatre by twice-weekly visits from aviation CRM trainers, who provided encouragement, coaching and feedback during a 3-month "bedding in" period. Typically, trainers observed a briefing, operation and debriefing, and then gave their own feedback to the team and to individuals. They refrained, however, from intervening during these processes, so as to allow local ownership to develop and minimise potential sensitivity among senior staff.

### EFFECTS OF CHANGE

The results of measures of attitudes, knowledge, behaviour, technical error and outcomes before intervention are shown in table 3.

### Effect of training on safety climate and non-technical performance

Training increased the mean SAQ teamwork climate score ( $t = -2.95$ ,  $p = 0.007$ ), while other SAQ components were not significantly affected, in keeping with expectations (see above). Team non-technical skills scores were significantly increased by training ( $t = -2.35$ ,  $p = 0.021$ ). LC teams experienced a marked increase in non-technical skills scores after training, but CEA teams did not (table 3). The effects of training were heterogeneous (table 4). Nursing subteam scores increased significantly, whereas anaesthetic and surgical subteam scores did not. Both P&D and T&C increased significantly, whereas SA and L&M did not. The non-technical skills scores of the independent third observer agreed very well with those of the two study observers ( $R_{wg} = 0.976$ ).

**Table 3** Comparison of main outcome measures before and after training

	All operations			Laparoscopic cholecystectomy			Carotid endarterectomy		
	Mean	SD	Test	Mean	SD	Test	Mean	SD	Test
SAQ teamwork climate									
Pre	64.1	18.9	$t = -2.95$	63.8	21.7	$t = -1.81$	64.5	15.0	$t = -2.34$
Post	69.2	18.2	$p = 0.007^*$	67.4	20.7	$p = 0.089$	72.0	14.2	$p = 0.041^*$
Total NOTECHS									
Pre	37.0	4.3	$t = -2.35$	35.5	4.9	$t = -3.32$	38.9	2.7	$t = 0.82$
Post	38.7	3.2	$p = 0.021^*$	38.7	2.9	$p = 0.001^*$	38.8	3.8	$p = 0.935$
Operating technical errors <sup>†</sup>									
Pre	1.73	1.5	$u = 1071.0$	2.6*	1.4	$u = 235.0$	0.7	0.7	$u = 206.5$
Post	0.98	0.9	$p = 0.009^*$	1.3	1.0	$p < 0.001^*$	0.4	0.6	$p = 0.236$
Non-operative procedural errors <sup>†</sup>									
Pre	8.48	4.2	$t = 4.383$	7.3	4.0	$t = 3.66$	9.9	4.1	$t = 2.42$
Post	5.16	3.7	$p < 0.001^*$	4.1	2.9	$p = 0.001^*$	7.0	4.1	$p = 0.020^*$
Operating time (min)									
Pre	103.7	38.9	$t = 0.29$	75.4	16.8	$t = 0.77$	137.1	29.8	$t = -1.28$
Post	101.2	49.1	$p = 0.770$	71.7	20.4	$p = 0.443$	151.1	42.5	$p = 0.208$
Length of stay (days)									
Pre	2.23	0.7	$u = 1263.5$	2.0	0.5	$u = 482.0$	2.5	0.7	$u = 196.5$
Post	2.02	0.6	$p = 0.086$	1.8	0.4	$p = 0.577$	2.3	0.7	$p = 0.192$

\*Significant difference at  $p < 0.05$ . †Figures represent mean values per operation. NOTECH, Non-Technical Skills; SAQ, Safety Attitudes Questionnaire.

**Table 4** Non-technical performance pretraining (n = 48) and post-training (n = 62)

Group statistics	NOTECHS*	SD	t	p Value
Dimension				
Leadership and management				
Pre	9.58	1.38	-0.70	0.485
Post	9.74	0.99		
Teamwork and cooperation				
Pre	9.33	1.31	-2.30	0.023*
Post	9.89	1.20		
Problem-solving and decision-making				
Pre	9.13	1.23	-3.17	0.002*
Post	9.76	0.86		
Situation awareness				
Pre	9.00	1.56	-1.29	0.199
Post	9.35	1.32		
Subteam				
Surgeons total				
Pre	13.54	1.70	-0.41	0.686
Post	13.66	1.39		
Anaesthetists total				
Pre	11.88	2.44	-0.29	0.772
Post	11.98	1.46		
Nursing total				
Pre	11.63	2.22	-2.16	0.033*
Post	12.47	1.86		

\*With this methodology, the total NOTECHS score possible was 12 for the team score on a single dimension and 16 for subteam scores.  
NOTECH, Non-Technical Skills scale.

### Effect of training on technical performance

Mean overall technical error rates declined significantly from 1.73 to 0.98 per operation ( $p = 0.009$ ). The change was more marked for LC, where it achieved significance, than for CEA, where it did not (see above comments). Non-operative procedural errors (NOPEs) were reduced by about 40% after training, which was highly significant for both operations (see table 3). The number of NOPEs per operation was three and 10 times higher than the number of operative technical errors in LC and CEA, respectively.

### Interactions between non-technical performance and technical errors

There was a significant but relatively weak association between overall non-technical skills score and operative technical error rate ( $\rho = -0.215$ ,  $p = 0.024$ ). Surgical subteam non-technical skills score ( $\rho = -0.236$ ,  $p = 0.013$ ), whole-team SA ( $\rho = -0.300$ ,  $p = 0.001$ ) and in particular surgical SA ( $\rho = -0.436$ ,  $p < 0.0001$ ) were much more strongly associated with technical error rates than overall team non-technical skills score.

**Table 5** Threats to validity

Threat to validity	Response/observation
Temporal effects	None identified with positive bias
Observer bias (NOTECHS)	Independent (uninvolved) third observer
Observer bias (OTE/NOPEs)	Objective, closely defined scales
Case selection	Consecutive series with specified exclusions
Selection bias in staff training	All relevant staff invited by same methods
Validity of scoring systems	Validation work (NOTECHS): previously validated scale (OCHRA)

NOPE, non-operative procedural error; NOTECH, Non-Technical Skills scale; OTE, operative technical error.

### Effect on clinical outcomes

The mean operating time was unaffected by the training (pretraining 103.7 (38.9) min, post-training 101.2 (49.1) min), and the mean LOS was not significantly reduced (pretraining 2.23 (0.7) days, post-training 2.02 (0.6) days).

Five complications were recorded from procedures performed before training, and three from those performed afterwards. One critical incident report was filed in the post-training group. No patient in either group returned to theatre or went to intensive treatment unit after surgery.

### NEXT STEPS

Training in non-technical skills resulted in improvement in attitudes to safety, team non-technical performance and technical error rates both in the operative field and outside it. Reductions in hospital stay and complications were not significant, which was unsurprising given the small study numbers. Previous observational studies have shown associations between teamwork, performance and safety in theatre,<sup>13 19</sup> but this is the first study to show improved technical outcomes after an intervention to improve non-technical skills. The results suggest that team training can improve the performance of theatre teams in a manner which is likely to bring clinical and operational benefits. Our findings are consistent with the MEDTEAMS study of ER staff,<sup>17</sup> which reported a significant reduction in clinical errors after team skills training. Further work is, however, required before this type of training can be generally recommended.

The before-after study design we used is at risk from time-related effects on the outcomes of interest, and our study population, while large by comparison with most studies of this type,<sup>27 28</sup> is small in absolute terms. We could not identify any external factors which changed during the study period and might have produced improvements in performance. We considered this possibility in terms of (a) personnel changes,

**Box 3 Case study 2: surgeon and anaesthetist**

A consultant surgeon was carrying out a carotid endarterectomy under local anaesthetic with a consultant anaesthetist giving sedation. The surgeon became concerned because the anaesthetist left the theatre several times when they were needed. This concern was apparent to other theatre staff but was not directly raised with the anaesthetist. After satisfactorily completing the operation, a formal debriefing was not conducted. The surgeon felt unable to raise the issue with the anaesthetist, for fear of upsetting his colleague, but he undertook to raise the issue the next time they worked together. A few weeks later, a similar incident was observed, and once again no debriefing or other communication took place.

(b) staffing levels and mix, (c) workload and case-mix, (d) staff morale factors, (e) equipment and resource problems. Factors (c) and (e) did not change during the study. There were few changes in the senior personnel involved. The hospital suffered a financial crisis during the study, which led to a number of staff redundancies, some reorganisation within the theatre teams and a decrease in staff morale. We were not able to train all the appropriate staff members, and a greater effect might have been achieved if training had been universal. Since the observers could not be blinded about which procedures were carried out by CRM trained team, and understood the study hypothesis, there was a risk of observer bias. Observer bias in the Oxford NOTECHS non-technical skills scoring method, the validation of which has been reported elsewhere,<sup>21</sup> was controlled for using an independent third observer. The methods used to record errors (operative technical errors and non-operative procedural errors) were semiobjective but highly structured. We cannot exclude a contribution of the “Hawthorne effect” of close observation on performance but would have expected such an effect to be maximal early in the observation period, that is prior to the intervention, since these effects typically fade with time. A summary of the threats to validity in the study and our responses to them is shown in table 5.

Overall, our results support the hypothesis that improving teamwork can improve technical performance, but some details were surprising. We found large decreases in technical error rates associated with apparently modest improvements in non-technical skills, and in the case of CEA a reduction in both technical error and non-operative procedural errors with no change in non-technical skills score. The training programme caused improvements mainly in teamwork and cooperation, problem-solving and decision-making, whereas situation awareness, particularly within the surgical team, was most strongly correlated with operative technical error and did not improve significantly. One interpretation of these findings is that surgical technical performance may have been affected indirectly by the better working environment resulting from improvements in nursing teamwork performance.

We cannot tell whether the NOTECHS non-technical skills scale reflects differences between teams equally well at all points on a scale between worst and best possible performance. Small changes in non-technical skills dimensions might therefore be associated with major improvements in technical performance at one or other end of the Oxford NOTECHS scale. It was interesting to consider why the training was more successful in the LC teams than the CEA team. The CEA team was much smaller and more stable than the LC teams and had a

much higher level of baseline non-technical performance. Our subjective impression, supporting the non-technical skills scores, was that the intervention was generally welcomed by nursing staff, but resisted or complied with passively and reluctantly by some senior medical staff, particularly in the CEA team. We postulate that increases in nursing non-technical skills for this team were cancelled out by decreases in the other subteams, but that an improvement in team performance nevertheless occurred, leading to a reduction in technical errors. Whole-team non-technical skills scoring as used in this study may lack sensitivity for those factors most strongly associated with improvements in technical errors, and may therefore require further refinement. Surgical team non-technical skills scores were highly reproducible and generally high, probably because the continuous activity of the team members made inferences about the cognitive aspects of their scores easier. By contrast, it was sometimes difficult to determine whether inactivity or silence in other subteams represented appropriate behaviour, loss of situational awareness or even loss of cooperation. We observed numerous instances which illustrate the need for improved communication, and the difficulty of accepting an open learning approach in the current surgical culture (see boxes 2–4). The power of hierarchy in inhibiting free communication was made particularly evident by observed examples of this type.

Our results are necessarily specific for one training programme, and further work will be needed to determine the nature and content of the most effective training intervention for reducing errors in theatre. We were impressed by the value of the “bedding-in” period, which allowed willing teams to develop competence and confidence by live practice with expert mentorship. Initially, the simple presence of the mentors was often sufficient to stimulate post-training teams to begin with a briefing, for example, whereas teams with no mentor present were likely to forget. We recommend that future training programmes incorporate prolonged on-the-job support. The aspects of the training which produced most subjective evidence of discomfort and difficulty were open honest debriefing and the use of a graduated challenge method to transmit urgent situational awareness information to authority figures. These appear to run counter to the prevailing professional culture, and are unlikely to be sustained without continuing direct efforts to promote them, as well as a general safety strategy designed to promote culture change. We suspect this will require strong, well-publicised institutional support and action at multiple levels, including education of new trainees, development of “champions” and substantial commitment of institutional resources to the development of official policies, incentives and regular retraining.<sup>29–31</sup> Anecdotal evidence from aviation experience suggests that management action to ensure compliance via policies and procedures may be necessary to ensure

**Box 4 Case study 3: hierarchy**

During the training, an experienced registrar was insistent that he had no problem in communicating with consultants. A few weeks later, he was observed in theatre assisting a complex operation which was not going well. His frustration and irritation were evident, but he said nothing until questioned afterwards by the observer. He felt that the difficulties were because the Consultant had adopted the wrong strategy, admitting “I didn’t say anything, did I?”

sustainable adoption in the face of cultural resistance. Such comprehensive action could yield further improvements in safety in the context of a formal quality-improvement programme. A larger trial with a randomised design and parallel explanatory qualitative analyses is needed to determine whether our findings can be reproduced in different settings, and whether clinical benefit can be directly demonstrated.

**Acknowledgements:** We would like to thank the management and the theatre, anaesthetic and surgical staff at the Oxford Radcliffe Hospital Trust for their permission to conduct this study, and for their participation in it.

**Funding:** This study was supported through a project grant from the BUPA Foundation, with additional funding from the Oxford Hospitals Medical Research Fund.

**Competing interests:** None.

**Ethics approval:** Ethics approval was provided by the Milton Keynes LREC (ref no 04/Q1603/35).

PM is the guarantor of the paper. He had the original idea and developed the study design together with KC and AM. AM and KC collected the data, and KC was principally responsible for its analysis. All authors contributed to the design of the intervention. P Smith acted as independent third observer for NOTECHS scores. KC began to develop the Oxford NOTECHS and NOPEs measurement methods while working at Great Ormond Street together with TD and GH from 2003, and these were further developed by the study group. AM was trained by the pilots in the observation of non-technical skills. PM wrote the first and final drafts of the paper, to which all authors contributed.

## REFERENCES

- Brennan TA, Leape LL. Adverse events, negligence in hospitalized patients: results from the Harvard Medical Practice Study. *Perspect Healthc Risk Manage* 1991;**11**:2–8.
- Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: Preliminary retrospective record review. *BMJ* 2001;**322**:517–19.
- Wolff AM, Bourke J. Reducing medical errors: a practical guide. *Med J Aust* 2000;**173**:247–51.
- Kohn LT, Corrigan JM, Donaldson MS. *To err is human: building a safer health system*. Washington, DC: National Academy Press, 2000.
- Leape LL. Error in medicine. *JAMA* 1994;**272**:1851–7.
- Rosenstein AH, O'Daniel M. Impact and implications of disruptive behavior in the perioperative arena. *J Am Coll Surg* 2006;**203**:96–105.
- Greenberg CC, Regenbogen SE, Studdert DM, et al. Patterns of communication breakdowns resulting in injury to surgical patients. *J Am Coll Surg* 2007;**204**:533–40.
- Kennedy I. Learning from Bristol: the report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–95. Command Paper: CM 5207. 2001.
- Flin R, O'Connor P, Mearns K. Crew resource management: improving team work in high reliability industries. *Team Perf Manage* 2002;**8**:68–78.
- Grogan EL, Stiles RA, France DJ, et al. The impact of aviation-based teamwork training on the attitudes of health-care professionals. *J Am Coll Surg* 2004;**199**:843–8.
- Haller G, Garnerin P, Morales MA, et al. Effect of crew resource management training in a multidisciplinary obstetrical setting. *Int J Qual Health Care* 2008;**20**:254–63.
- Catchpole KR, Giddings AE, de Leval MR, et al. Identification of systems failures in successful paediatric cardiac surgery. *Ergonomics* 2006;**49**:567–88.
- Carthey J, de Leval MR, Wright DJ, et al. Behavioural markers of surgical excellence. *Saf Sci* 2003;**41**:409–25.
- Mishra A, Catchpole K, Dale T, et al. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surg Endosc* 2008;**22**:68–73.
- Healey AN, Undre S, Vincent CA. Defining the technical skills of teamwork in surgery. *Qual Saf Health Care* 2006;**15**:231–4.
- Yule S, Flin R, Paterson-Brown S, et al. Development of a rating system for surgeons' non-technical skills. *Med Educ* 2006;**40**:1098–104.
- Guerlain S, Adams RB, Turrentine FB, 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.
- Lingard L, Reznick R, Espin S, et al. Team communications in the operating room: talk patterns, sites of tension, and implications for novices. *Acad Med* 2002;**77**:232–7.
- Catchpole K, Giddings AE, Wilkinson M, et al. Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007;**142**:102–10.
- Morey JC, Simon R, Jay GD, 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**:1553–81.
- 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*. In press.
- Flin RH, Martin L, Goeters K-M, et al. Development of the NOTECHS (Non-Technical Skills) system for assessing pilots' CRM skills. *Hum Perf Extreme Environ* 2003;**3**:95–117.
- Joice P, Hanna GB, Cuschieri A. Errors enacted during endoscopic surgery—a human reliability analysis. *Appl Ergon* 1998;**29**:409–14.
- Tang B, Hanna GB, Joice P, et al. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Arch Surg* 2004;**139**:1215–20.
- Sexton JB, Helmreich RL, Neilands TB, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;**6**:44.
- Swinscow TDV, Campbell MJ. *Statistics at square one*. 9th edn. London: BMJ Books, 1997.
- Volpe CE, Cannon-Bowers JA, Salas E, et al. The impact of cross-training on team functioning: an empirical investigation. *Hum Factors* 1996;**38**:87–100.
- Lingard L, Espin S, Rubin B, et al. Getting teams to talk: development and pilot implementation of a checklist to promote interprofessional communication in the OR. *Qual Saf Health Care* 2005;**14**:340–6.
- Helmreich RL, Merritt AC. *Culture at work in aviation and medicine*. Aldershot: Ashgate, 1998.
- Reason J. *Managing the risks of organisational accidents*. Aldershot: Ashgate, 1997.
- Wilson KA, Burke CS, Priest HA, et al. Promoting health care safety through training high reliability teams. *Qual Saf Health Care* 2005;**14**:303–9.