The Impact of Nontechnical Skills on Technical Performance in Surgery: A Systematic Review

Louise Hull, MSc, Sonal Arora, PhD, MRCS, Rajesh Aggarwal, PhD, FRCS, Ara Darzi, MD, FACS, Charles Vincent, PhD, Nick Sevdalis, PhD

BACKGROUND: Failures in nontechnical and teamwork skills frequently lie at the heart of harm and near-misses in the operating room (OR). The purpose of this systematic review was to assess the impact of nontechnical skills on technical performance in surgery.

STUDY DESIGN: MEDLINE, EMBASE, PsycINFO databases were searched, and 2,041 articles were identified. After limits were applied, 341 articles were retrieved for evaluation. Of these, 28 articles were accepted for this review. Data were extracted from the articles regarding sample population, study design and setting, measures of nontechnical skills and technical performance, study findings, and limitations.

RESULTS: Of the 28 articles that met inclusion criteria, 21 articles assessed the impact of surgeons’ nontechnical skills on their technical performance. The evidence suggests that receiving feedback and effectively coping with stressful events in the OR has a beneficial impact on certain aspects of technical performance. Conversely, increased levels of fatigue are associated with detriments to surgical skill. One article assessed the impact of anesthesiologists’ nontechnical skills on anesthetic technical performance, finding a strong positive correlation between the 2 skill sets. Finally, 6 articles assessed the impact of multiple nontechnical skills of the entire OR team on surgical performance. A strong relationship between teamwork failure and technical error was empirically demonstrated in these studies.

CONCLUSIONS: Evidence suggests that certain nontechnical aspects of performance can enhance or, if lacking, contribute to deterioration of surgeons’ technical performance. The precise extent of this effect remains to be elucidated. (J Am Coll Surg 2012;214:214–230. © 2012 by the American College of Surgeons)
however, this hypothesis remains to be empirically tested. It is also unclear exactly which nontechnical skills have the most significant impact on technical outcomes: the assumption that they all affect performance and safety in an equal manner remains to be investigated. The fact that nontechnical skills and technical performance are considered and consequently explored as 2 independent skills, is reflected both in the evidence base and in current curricula. This is highlighted by the separate reviews of each skill set and by training programs that focus exclusively on 1 domain to the neglect of the other. What is required is a clearer understanding of the complex and interdependent relationships between nontechnical skills and technical performance that more accurately represent the clinical realities of working in the OR. Without this, effective training interventions will remain elusive. In contrast, a detailed mapping exercise could allow patient safety efforts to focus on interventions that target the nontechnical skills most likely to have an impact on technical outcomes.

Not surprisingly therefore, the integration of nontechnical skills into mainstream surgical education remains limited to a few programs such as the American College of Surgeons-Association of Program Directors in Surgery (ACS-APDS) Phase 3 curriculum and Team Strategies and Tools to Enhance Performance and Patient Safety (Team STEPPS). A poor understanding of the relevance of nontechnical skills to technical performance could offer an explanation. Conflicting results on the impact of nontechnical skills, produced by individual studies with variable quality, have not assuaged these concerns. A synthesis of the existing evidence on both skill sets could highlight the importance of nontechnical skills to the clinical community and therefore drive an evolution of surgical curricula to encompass training and feedback on these skills, both in the OR and in simulation-based settings.

The aim of this systematic review was to synthesize the existing literature on the impact of nontechnical skills on technical performance in surgery. Specifically, the primary aim was to evaluate the impact of the nontechnical skill(s) of each surgical team-member (surgeons, anesthesiologists, and OR nurses) on their technical performance in the OR. A secondary aim was to identify assessment tools used to assess nontechnical skills and performance.

**METHODS**

**Data sources**

Databases searched included MEDLINE (OVID) (1980-week 2, April 2010), EMBASE (OVID) (1980-week 2, April 2010), and PsycINFO (OVID) (1987-week 2, April 2010). The Cochrane Database of Systematic Reviews was also searched. MeSH terms were identified to ensure the search was comprehensive. The search strategy is detailed in the Appendix. After combining categories A, B, and C, using the Boolean term “AND”, the following limits were applied: publication date: 1980-week 2, April 2010; English; and Humans. The last search was conducted on April 26, 2010. The grey literature (unpublished studies, with limited distribution, eg, conference abstracts) and reference lists of the included articles were also searched for additional citations. This resulted in the identification of 3 additional articles that warranted full review.

Retrieved citations were assessed by the primary reviewer with a background in psychology (LH), who screened the titles and abstracts to identify relevant studies based on predefined inclusion criteria, as follows:

1. Data on the assessment of technical performance of surgeons, anesthesiologists, or OR nurses

2. Data on the assessment of nontechnical skill(s) of surgeons, anesthesiologists, or OR nurses and/or study designs that include factors known to affect performance (eg, studies that create stressful conditions in a simulated OR and assess their impact on technical performance)

3. Direct empirical evaluation of the impact of nontechnical skills on technical performance.

These criteria were designed to identify studies that specifically captured the impact of nontechnical skills on technical performance in surgical teams.

A second reviewer with a background in surgery (SA) screened all abstracts to ensure reliability. Reliability in screening between the 2 reviewers was assessed using Cohen’s kappa at the abstract and full-text review phases. Any discrepancies were resolved by consensus. Data were extracted from included articles using a structured data abstraction form to ensure articles were evaluated in a consistent manner (the form is available upon request). Using this form, information regarding study design and setting, specialty and sample size, nontechnical skill(s) measured and assessment tools, technical skill(s) measured and assessment tools, and key findings were extracted. A critical appraisal of each study was completed.

**RESULTS**

**Selected articles**

A flow diagram of the search results is illustrated in Figure 1. The search yielded 2,041 citations, of which 881 articles were excluded after limits were applied. The remaining 1,160 articles were retrieved for title evaluation. There were 819 articles excluded based on title evaluation. The remain-
Figure 1. Flow of articles in review.
ing 341 articles were retrieved for abstract evaluation. Two reviewers (LH and SA) independently screened all 341 abstracts for eligibility. Of these articles, 87 were judged to warrant full review, and 25 of these met inclusion criteria. Agreement between the 2 reviewers was high for both phases (abstract and full-text: kappa = 0.76, p < 0.001 and kappa = 0.90, p < 0.001, respectively). Three additional articles were identified via hand search, giving a total of 28 articles for this review.

Data synthesis
Article findings were categorized according to study populations (surgeons, anesthesiologists, OR nurses). Meta-analysis was inappropriate for the articles selected because of wide heterogeneity in research designs and outcomes measures used.

Study setting
Most studies were conducted in simulated environments (22 of 28). The type and fidelity of simulated studies varied; nine of 28 were laboratory-based studies that used bench-top models,15-23 3 of 28 were hybrid simulations (combining bench-top models and simulated patients),24-26 6 of 28 used virtual reality simulators,27-32 and 4 of 28 studies were full-team simulations conducted in simulated ORs.26,33-34 The remaining 6 studies were conducted in real ORs.35-40

Measures of nontechnical skills
In order to explore the impact of nontechnical skills on technical performance in surgery, it is essential that assessment tools with evidence of validity and reliability (ie, psychometrically robust) are used. However, the psychometric properties of the nontechnical skill assessment tools used in the 28 articles within this review varied greatly (Table 1); from assessment tools developed for use in a particular study (without validation or reliability testing),20,27,31,33 to tools designed to assess nontechnical skills specifically in surgical contexts that have received extensive psychometric testing.34-38, 41 In only 7 of the 28 included articles did the authors use predeveloped tools with evidence of validity and reliability to measure nontechnical performance: 4 used the Oxford Non-Technical Skill (NOTECHS) System,26,35-38 1 used the Observational Teamwork Assessment for Surgery (OTAS),34 1 used the Revised NOTECHS,26 and 1 used Anesthetists’ Non-Technical Skills (ANTS).41

In an additional 9 studies, nontechnical skills were not assessed directly; rather, these studies used research designs that assessed the impact of factors known to affect performance.15-19,21,23,25,42 For example, in 1 study by Backstein and colleagues,23 the impact of providing performance feedback on the technical performance of the participants was assessed. Seeking feedback has been defined as a nontechnical skill that can lead to performance enhancement.7 So although a nontechnical skill (feedback) was not directly assessed per se in reports such as this, such studies were still included in our review because manipulation of a key variable (ie, the nontechnical skill of feedback) allowed inferences to be drawn about its impact on technical performance. Finally, in 2 studies, an observer recorded descriptions of operative events that were subsequently categorized using human factors theory into teamwork failures.39,40

Measures of technical performance
Valid and reliable assessment of technical performance is also essential for robust conclusions about end outcomes. Technical performance measures varied greatly both across and within the reviewed studies. Seventeen studies in this review measured dexterity parameters. This included time to complete the task/operation,16,17,19,20,23,27-29,31,32,35,36,42 economy of motion,27,29,31 tool movement smoothness,28,30 instrument smoothness,19 hand movement,28,30 instrument path length,19,32 gesture proficiency,28 and hand motion efficiency.18,21 Twenty-five studies used quality of technical performance as an end outcome. Measures for this included number of errors,15,19,27,29,32,39,40 errors in technique assessed via OCHRA,35-38 and final product quality.7,8,25,32 Other assessment tools to capture quality of technical performance included global rating scales such as OSATS,18,21,23,33,42 and MOSAT.41 Checklists were also used in the form of task-specific checklists,15,22-25,33 essential-item checklist,33 procedure-specific skill,26 procedural problems and errors (NOPES),35,36 and anesthetic checklists (assessing medical knowledge and technical skill of anesthesiologists).41

The impact of surgeons’ nontechnical skills on their technical performance
As indicated in Table 2, 21 of the 28 reports assessed the impact of surgeons’ nontechnical skills on their technical performance.15-34,42 The majority of these studies (18 of 21) assessed the impact of a single nontechnical skill on technical performance; feedback (n = 9),15-19,21,23,25,42 stress (n = 4),22,29,32,34 fatigue (n = 4),27,28,30,31 and communication (n = 1)34 were the skills that were individually investigated. Three of the 21 reports measured the impact of multiple nontechnical skills on performance.20,26,33 Detailed findings are presented below.

Stress and technical performance
A substantial body of research on stress and human performance shows that although a certain amount of stress can
<table>
<thead>
<tr>
<th>NTS assessment Tool (Study)</th>
<th>NTS assessed</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
</table>
| Observation Teamwork Assessment for Surgery (OTAS) | 1. Communication  
2. Leadership  
3. Cooperation  
4. Coordination  
5. Team monitoring | Inter-rater reliability  
Construct valid  
Content valid |
| Oxford Non-Technical Skills (NOTECHS) | 1. Leadership and management  
2. Teamwork and communication  
3. Problem solving and decision making  
4. Situational awareness | Inter-rater reliability  
Predictive validity  
Concurrent valid  
Convergent valid |
| Anesthetists’ Non-Technical Skills (ANTS) | 1. Resource management  
2. Planning  
3. Leadership  
4. Communication | Inter-rater reliability  
Internal consistency  
Completeness validity  
Observability validity |
| Revised NOTECHS | 1. Cooperation  
2. Leadership and managerial skills  
3. Situational awareness and vigilance  
4. Decision making  
5. Communication and interaction | Internal consistency  
Inter-rater reliability  
Construct valid |
| Line Operations Safety Audit Checklist (LOSA; selected elements) | 1. Preoperative preparation  
2. Communication and interaction  
3. Vigilance/situational awareness  
4. Leadership | Inter-rater reliability  
No validity evidence reported |
| 5-point rating scale | Teamwork  
1. Communication  
2. Cooperation | No reliability evidence reported  
No validity evidence reported |
| Teamwork events characterized according to human factors theory | Teamwork failures/disruptions | Results categorized according to human factors theory based on consensus agreement between observer, cardiac surgeon, and human factors scientist  
No validity evidence reported |
| State Trait Anxiety Inventory (STAI) | Stress | Reliable  
Valid |
| 10-point scale | Stress | No reliability evidence reported  
Valid |
| Imperial Stress Assessment Tool (ISAT) | Stress  
1. State Trait Anxiety Inventory (STAI)  
2. Heart rate  
3. Salivary cortisol | STAI: Internal consistency  
Content valid  
Concurrent valid |
| SVF78 Stress-Coping Questionnaire | Stress coping | Internal consistency  
Valid |
| Communication-based Objective Structured Clinical Examination (OSCE) | Communication  
1. Empathy  
2. Coherence  
3. Verbal expression  
4. Nonverbal expression  
5. Overall | No reliability evidence reported  
Construct valid |
| Utterance frequency (UF) | Communication | No reliability evidence reported  
No validity evidence reported |
| 4-point subjective rating scale | Fatigue | No reliability evidence reported  
No validity evidence reported |
| 10-point rating scale | Fatigue | No reliability evidence reported  
No validity evidence reported |
| Behrenz Fatigue Questionnaire | Fatigue | No reliability evidence reported  
No validity evidence reported |

NTS, nontechnical skill.
Table 2. Characteristics of Studies Investigating the Impact of Surgeons’ Nontechnical Skills on their Technical Performance

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Subjects and speciality</th>
<th>Study design and setting</th>
<th>Nontechnical skill(s) assessed and assessment measure(s)</th>
<th>Technical skill(s) assessed and assessment measure(s)</th>
<th>Findings</th>
<th>Critical appraisal*</th>
</tr>
</thead>
</table>
| Arora        | 2010 | 18 junior surgeons      | Prospective experimental study; simulated setting | Stress ISAT | 1. Path length  
2. Time taken  
3. Number of errors MIST-VR simulator | Significant positive correlation between stress and economy of motion, number of errors, and time taken | Junior trainees completed relatively simple tasks under moderately stressful conditions. Small sample size. Simulated setting |
| Backstein    | 2004 | 29 surgical residents (PGY1-5) | Intervention study; simulated setting | Feedback manipulation 1. Control (no feedback)  
2. Experimental 1 (video and self-review)  
3. Experimental 2 (video and expert feedback) | 1. Task-specific checklist  
2. Time taken to complete procedure  
3. Global rating scale OSATS | No significant differences between the 3 different feedback conditions and technical performance | Small sample size. Combining experienced residents with junior residents. Outcomes measures may lack sensitivity to identify differences among more experienced surgeons. One feedback session may not have been sufficient to produce a change. Restricted number of residents in combination with the higher level of surgical proficiency at the outset of the study may have impacted the ability to obtain statistically significant findings. Simulated setting |
| Backstein    | 2005 | 26 surgical residents (PGY1) | Intervention study; simulated setting | Feedback manipulation 1. Experimental 1 (expert feedback)  
2. Experimental 2 (video feedback with expert review) | 1. Task-specific checklist MOSAT  
2. Global rating scale MOSAT | No significant differences between the 2 conditions and technical performance | Technical assessment scales may lack sensitivity needed to measure the subtle improvements in surgical skill. Single procedure. Small sample size. Both groups received some form of feedback (video/expert) which may not have been qualitatively different to identify differences in technical performance. Simulated setting |
| Black        | 2010 | 30 surgeons (10 junior, 10 senior trainees, and 10 consultants) | Prospective observational study; simulated setting | 1. Cooperation  
2. Leadership and managerial skills  
3. Situational awareness and vigilance  
4. Decision making  
5. Communication and interaction Revised NOTECHS | 1. Global rating scale OSATS  
2. Procedure specific-skill ICEPS for CEA | Significant positive correlation between technical and nontechnical performance in noncrisis and crisis scenario | Inability to blind the assessors to the seniority of trainee. Small sample size. Simulated setting. Single procedure |
| DeMaria      | 2005 | 17 surgeons (16 residents [PGY1-5] and 1 attending) | Prospective experimental study; simulated setting | Fatigue manipulation 1. Before night call  
2. After night call  
Sleep diary: fatigue measured on a 4-point scale | 1. Economy of motion  
2. Time to complete task  
3. Errors by each hand/foot  
4. Total time  
5. Total number of errors  
6. Overall score MIST-VR simulator | Economy of motion improved significantly post-call for dominant and non-dominant hands in 2/6 tasks  
Total time to complete task improved significantly post-call in 3/6 tasks  
Significantly fewer errors made post-call in 1/6 tasks  
Overall scores significantly improved for 4/6 tasks | Fatigue level: Pre-test: 82% participants felt “well-rested” or “tested”  
Post-test: 76.5% felt “not rested”  
30 residents took part in pre-call testing, only 16 were re-tested post-call. Small sample size. Homogenous sample. Simulated setting |
Table 2. Continued

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Subjects and speciality</th>
<th>Study design and setting</th>
<th>Nontechnical skill(s) assessed and assessment measure(s)</th>
<th>Technical skill(s) assessed and assessment measure(s)</th>
<th>Findings</th>
<th>Critical appraisal*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastridge</td>
<td>2003</td>
<td>31 surgical residents (PGY1-5)</td>
<td>Prospective experimental study; simulated setting</td>
<td>Fatigue manipulation 1. Pre-call (rested) 2. On-call (rested) 3. Post-call (acutely sleep deprived) Fatigue measured on a 10-point scale</td>
<td>1. Speed 2. Error 3. Economy of motion MIST-VR simulator</td>
<td>Number of errors and time to complete all 6 tasks increased post-call (fatigue state) Fatigue level: significant increase from pre-call to post-call</td>
<td>Homogenous sample. Clinical significance of findings in a simulator is arguable. Small sample size Simulated setting</td>
</tr>
<tr>
<td>Gerdes</td>
<td>2008</td>
<td>30 surgeons (5 residents [PGY3] and 9 attending)</td>
<td>Prospective experimental study; simulated setting</td>
<td>Fatigue manipulation 1. Pre-call 2. Post-call Behrenz Fatigue Questionnaire</td>
<td>1. Hand movement Cyberglove and Polhemus Limby Tracker 2. Tool movement 3. Cognitive error 4. Psychomotor error 5. Time to complete Sensable Haptic Joystick simulator</td>
<td>Significant decrement in proficiency measures post-call (fatigue state) No significant decrement in number of psychomotor errors post-call (fatigue state) Significant increase in cognitive errors post-call (fatigue state) Increased fatigue ratings associated with increased errors ($R^2 = 0.92$)</td>
<td>Small sample size Simulated setting</td>
</tr>
<tr>
<td>Hassan</td>
<td>2006</td>
<td>24 surgeons (12 final-year medical students and 12 junior residents [PGY1-3])</td>
<td>Prospective observational study; simulated setting</td>
<td>Stress coping strategies SYF78 Stress-Coping Questionnaire</td>
<td>1. Time to complete task 2. Number of errors 3. Economy of motion LapSim simulator</td>
<td>Task difficulty: Easy Negative stress coping correlated positively with time to complete task Task difficulty: Difficult Negative stress-coping strategies positively correlated with time to complete task and number of errors Borderline significant correlation positive correlation between negative stress-coping and economy of motion</td>
<td>Homogenous sample Simulated setting Self-report stress-coping questionnaire. Small sample size</td>
</tr>
<tr>
<td>Jensen</td>
<td>2009</td>
<td>45 surgical residents (PGY1-2)</td>
<td>Intervention study; simulated setting</td>
<td>Feedback manipulation 1. Control (no feedback) 2. Experimental (expert feedback)</td>
<td>1. Global rating scale OSATS 2. Time to complete task 3. Final product quality (esthetic rating scale for skin closure and anastomotic leak pressure for bowel anastomosis)</td>
<td>No significant difference between the 2 groups in global rating scale and time to complete procedure No significant difference between the 2 groups for esthetic score for skin closure Difference in anastomotic leak pressure: expert feedback group displayed superior performance</td>
<td>Validity of OSATS not studied formally for use with a video-recorded performance. The skin esthetic rating scale may not be able to discriminate at a fine enough level to show a difference. Small sample size and single instructor limits generalizability Simulated setting</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Study design and setting</th>
<th>Nontechnical skill(s) assessed and assessment measure(s)</th>
<th>Technical skill(s) assessed and assessment measure(s)</th>
<th>Findings</th>
<th>Critical appraisal*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeBlanc</td>
<td>2008</td>
<td>Prospective experimental study; simulated setting</td>
<td>Stress manipulation (1. Pre-exam [low stress], 2. Examination [high stress], 3. Post-exam [low stress] Stress measured on a 10-point scale)</td>
<td>1. Global rating scale OSATS 2. Task-specific checklist</td>
<td>Significantly better performance in high stress condition of task-specific checklist scale No significant difference in performance (low vs high stress) on global rating scale Stress levels significantly higher high stress vs low stress condition</td>
<td>Subjective measure of stress, Small sample size, Specific stress manipulation: socioevaluative Simulated setting</td>
</tr>
<tr>
<td>First author</td>
<td>Year</td>
<td>Subjects and speciality</td>
<td>Study design and setting</td>
<td>Nontechnical skill(s) assessed and assessment measure(s)</td>
<td>Technical skill(s) assessed and assessment measure(s)</td>
<td>Findings</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>-------------------------</td>
<td>--------------------------</td>
<td>----------------------------------------------------------</td>
<td>------------------------------------------------------</td>
<td>----------</td>
</tr>
</tbody>
</table>
| Moulton      | 2009 | 32 surgical trainees (16 medical students [4th y], 16 residents [PGY1]) | Intervention study; simulated setting | Feedback manipulation 1. Control (no feedback) 2. Experimental (expert feedback) | 1. Global rating scale OSATS 2. Task-specific checklist | No significant difference between the 2 groups in technical performance | Small sample size  
Simulated setting  
Received feedback on communication performance ONLY  
Surgeon-patient interaction |
Homogeneous sample  
Medical students with no previous laparoscopic experience  
Simple task (low technical difficulty)  
Simulated setting |
| Porte        | 2007 | 45 surgical trainees (1st y medical students) | Intervention study; simulated setting | Feedback manipulation 1. Motion analysis feedback (no criterion) 2. Motion analysis feedback (criterion) 3. Expert feedback | 1. Global rating scale 2. Hand motion efficiency ICSAD | Statistically significant improvement in global rating scale and ICSAD scores for all 3 groups from pre-post test Groups 1 and 2 displayed no significant improvement between pre-delayed post-test scores. Group 3 showed sustained improvement (pre-delayed post-test) | Homogeneous sample  
Simple task (low technical difficulty)  
Simulated setting |
| Rogers       | 1998 | 82 surgical trainees (junior/senior medical students) | Intervention study; simulated setting | Feedback manipulation 1. Control (no feedback) 2. Experimental (expert feedback) | 1. Knot square 2. Quality of knot tying 3. Time taken to complete Not specified | No significant difference between the 2 groups in knot square and time to complete  
Significant difference between the 2 groups in quality of knot tying; feedback group superior performance | Homogenous sample  
Simple task (low technical difficulty)  
Simulated setting |
| Rogers       | 2000 | 105 surgical trainees (junior/senior medical students) | Intervention study; simulated setting | Feedback manipulation 1. Control (no feedback) 2. Experimental (expert feedback) | 1. Knot square 2. Quality of knot tying 3. Time taken to complete Not specified | No significant difference between the 2 groups in knot square and time to complete  
Significant difference between the 2 groups in quality of knot tying; feedback group superior performance | Homogenous sample  
Simple task (low technical difficulty)  
Simulated setting |

(Continued)
<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Subjects and speciality</th>
<th>Study design and setting</th>
<th>Nontechnical skill(s) assessed and assessment measure(s)</th>
<th>Technical skill(s) assessed and assessment measure(s)</th>
<th>Findings</th>
<th>Critical appraisal*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xeroulis</td>
<td>2007</td>
<td>60 surgical trainees (1st year medical students)</td>
<td>Intervention study; simulated setting</td>
<td>Feedback manipulation 1. Control (no feedback) 2. Self-study with CBVI 3. Expert feedback (concurrent feedback) 4. Expert feedback (summary feedback)</td>
<td>1. Global rating scale OSATS 2. Hand motion efficiency ICSAD</td>
<td>Pre-test global rating scale: No significant differences between all groups. Post-test global rating scale: All groups improved from pre to post-test. All 3 experimental groups displayed superior performance compared to control group. 1-month post-test retention global rating scale: CBVI and summary feedback groups retained superior performance compared to controls. Post-test hand motion efficiency: no significant differences between all groups</td>
<td>Does not address utility of these methods (feedback) in more complex tasks and whether these skills are transferable into clinical setting. Homogenous sample</td>
</tr>
<tr>
<td>Zhengl</td>
<td>2008</td>
<td>44 surgical trainees (8 medical students/office staff, 13 PGY1-3, 10 PGY4-5, 7 fellows and 6 attending)</td>
<td>Prospective observational study; simulated setting</td>
<td>Teamwork 1. Communication 2. Cooperation 5-point rating scale.</td>
<td>1. Speed 2. Accuracy LISETT</td>
<td>Positive correlation between LISETT score and team quality score Top self-rated teams performed the LISETT task significantly better than the lowest self-rated teams</td>
<td>Self-evaluation of team quality was affected by the status of the surgeons within a team. Office staff included in sample. Self-rated team quality NTS assessment measure not validated</td>
</tr>
</tbody>
</table>

CBVI, computer-based video instruction; EPA, End Product Assessment; HR, Heart rate; HRV, Heart rate variability; ICEPS for CEA, Imperial College Evaluation of Procedure-specific Skill scale for carotid endarterectomy; ICSAD, Imperial College Surgical Assessment Device; ISAT, Imperial Stress Assessment Tool; KP, knowledge of performance; KR, knowledge of result; LISETT, The Legacy Inanimate Systems for Endoscopic Team Training; LOSA, Line Operations Safety Audit Checklist; MIST-VR, Minimally Invasive Surgical Trainer-Virtual Reality; MOSAT, Mini Objective Structured Assessment of Technical Skills; NTS, nontechnical skills; OSCE, Objective Structured Clinical Evaluation; OTAS, Observational Teamwork Assessment for Surgery; Revised NOTECHS, Revised Non-Technical Skills; OSATS, Objective Assessment of Technical Skills; PGY, postgraduate year; SC, Salivary Cortisol; STAI, State Trait Anxiety Inventory.

*Italics denote author critical appraisal.
improve performance by enhancing concentration and focus, excessive stress compromises performance.\textsuperscript{43} Despite the fact that coping with stress is an important nontechnical skill,\textsuperscript{7} as determined from a recent systematic review in this field, research on stress within surgery is sparse.\textsuperscript{44} This review identified 4 studies.\textsuperscript{22,29,32,34} Arora and associates\textsuperscript{32} found that increased stress due to inexperience and unfamiliarity with a task was related to poorer technical performance. In contrast, a study assessing the impact of examination stress found that moderate increases in stress enhanced residents’ performance on a simulated task.\textsuperscript{22} The remaining 2 studies focused on coping with stress.\textsuperscript{29,34} The first study found that negative stress—coping strategies were associated with poorer laparoscopic performance on a virtual reality simulator;\textsuperscript{29} the second found that enhanced coping strategies, even with multiple stressors, significantly improved the quality of the operative end product.\textsuperscript{34}

**Performance feedback and technical performance**

Seeking advice and feedback is a critical intraoperative nontechnical skill.\textsuperscript{7} Nine studies assessed the impact of feedback on performance.\textsuperscript{15-19,21,23,25,42} Of these, 6 studies\textsuperscript{16-19,21,42} found that receiving feedback from an expert enhanced technical performance with regard to its overall quality,\textsuperscript{18,21} economy of motion,\textsuperscript{18,19} number of errors,\textsuperscript{19} time taken to complete task,\textsuperscript{19} and overall quality of end product.\textsuperscript{16,17,21,42} However, this positive effect of feedback did not equally affect all aspects of technical performance in these studies; some performance parameters improved, but others did not.\textsuperscript{16,17,21,42} Moreover, 3 other studies\textsuperscript{15,23,25} found no improvement in technical skill in surgeons who received feedback compared with those who did not.

**Fatigue and technical performance**

Although safe management of fatigue is considered a key nontechnical skill,\textsuperscript{7} most studies investigating surgeon fatigue have not included technical performance as an endpoint. Of the 28 studies forming the basis of this review, 4 explicitly assessed the impact of fatigue (caused by sleep deprivation) on technical performance.\textsuperscript{27,28,30,31} Three of these studies\textsuperscript{28,30,31} found that increased levels of fatigue were associated with more technical errors,\textsuperscript{31} time to complete the task,\textsuperscript{31} and instrument handling.\textsuperscript{28,30} However, this negative effect of fatigue did not equally affect all aspects of technical performance; some performance parameters deteriorated, but others did not.\textsuperscript{28,30,31} In 1 study, however, DeMaria and coworkers\textsuperscript{27} found that for certain tasks on the Minimally Invasive Surgical Trainer-Virtual Reality (MIST-VR) simulator, fatigue actually enhanced economy of motion and resulted in faster task completion with fewer errors.

**Communication and technical performance**

Only 1 study\textsuperscript{24} specifically assessed the impact of surgeons’ communication skills on their technical performance and found no correlation between them. However, this study assessed surgeons’ ability to communicate effectively with the patient, rather than with their OR team.

**The combined impact of multiple nontechnical skills and technical performance**

Concurrent assessment of multiple nontechnical skills is important because safe surgery requires an interplay of these skills. Only 3 studies\textsuperscript{20,26,33} concurrently assessed more than 1 nontechnical skill. Black and colleagues\textsuperscript{26} assessed nontechnical skills using Revised NOTECHS\textsuperscript{45} (communication and interaction, situation awareness and vigilance, cooperation and team skills, leadership and managerial skills, and decision making) as well as the technical performance of surgeons performing a carotid endarterectomy in simulated routine and crisis scenarios. In both scenarios, a strong positive correlation (routine: $r = 0.80$, $p < 0.001$; crisis: $r = 0.85$, $p < 0.001$) between overall technical and nontechnical performance was found. Unfortunately, more detailed correlational analyses between specific nontechnical skills and technical performance were not reported. Mirroring this, another study also found that high levels of team communication and cooperation were positively correlated ($r = 0.39$, $p < 0.008$) with the speed and accuracy of surgeons’ performing peg transportation and intracorporeal suture simulated tasks.\textsuperscript{20} Moorthy and associates\textsuperscript{35} also measured preoperative preparation, communication and interaction, vigilance and situational awareness, and leadership in surgeons performing a saphenofemoral junction high tie procedure. In this study, the relationship between overall technical and nontechnical performance was not significant ($\rho = 0.24$, $p = 0.23$); correlational analyses between each nontechnical skill and technical performance were not reported.

**The impact of OR team members’ nontechnical skills on surgeons’ technical performance**

As shown in Table 3, 6 reports assessed the impact of the entire OR team’s (surgical, anesthetic, and nursing members) nontechnical skills on surgical technical performance.\textsuperscript{35-40} Four reports (based on 2 datasets) assessed nontechnical skills using the Oxford NOTECHS System. Collectively these studies found weak negative correlations ($r_s = -0.16$ to $-0.27$) between overall team nontechnical performance and technical errors made by the operating surgeon.\textsuperscript{25-38} The nontechnical skill that emerged as more relevant in these reports was situational awareness, such that better situational awareness was associated with fewer surgical errors ($r = -0.72$, $p < 0.001$, $F(2,42) = 7.93$, $p < 0.001$).\textsuperscript{35} Moreover, higher-
### Table 3. Characteristics of Studies Investigating the Impact of Operating Room Team Members’ Nontechnical Skills on Surgeons’ Technical Performance

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Subjects and speciality</th>
<th>Study design and setting</th>
<th>Nontechnical skill(s) assessed and assessment measure(s)</th>
<th>Technical skill(s) assessed and assessment measure(s)</th>
<th>Findings</th>
<th>Critical appraisal*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catchpole McCulloch</td>
<td>2008&lt;sup&gt;10&lt;/sup&gt;</td>
<td>103 OR teams (surgeons, nurses, and anesthesiologists; 48 preintervention and 55 postintervention)</td>
<td>Intervention study; real OR</td>
<td>1. LM 2. TC 3. PD 4. SA Oxford NOTECHS</td>
<td>1. Errors in surgical technique OCHRA 2. Procedural problems and errors NOPEs 3. Operative time</td>
<td>Preintervention: Relationship between technical error and surgical SA Operative time significantly affected by surgical LM Operative time significant interaction between operative type and anesthetic LM Procedural problems and errors affected by nursing LM Postintervention: Significant weak negative correlation between overall NOTECHS scores and surgical technical errors Significant weak negative correlation between entire team SA, and moderate negative correlation between surgical SA, and technical error</td>
<td>Single site study Small sample size in absolute terms.</td>
</tr>
<tr>
<td>ElBardissi Wiegmann</td>
<td>2008&lt;sup&gt;8&lt;/sup&gt;</td>
<td>31 surgical cases; 5 surgeons (nurses, anesthesiologists and perfusionist)</td>
<td>Prospective observational study; Real OR</td>
<td>Teamwork failures/disruptions</td>
<td>1. Surgical errors (written documentation via observation)</td>
<td>Strong positive correlation between teamwork disruptions and surgical errors</td>
<td>Single site study Small sample size (surgeons n = 5) Nonvalidated assessment measures</td>
</tr>
<tr>
<td>Mishra</td>
<td>2008&lt;sup&gt;5&lt;/sup&gt;</td>
<td>65 OR teams (surgeons, nurses, and anesthesiologists; 26 preintervention, 39 postintervention)</td>
<td>Intervention study; real OR</td>
<td>1. LM 2. TC 3. PD 4. SA Oxford NOTECHS</td>
<td>1. Errors in surgical technique OCHRA</td>
<td>Preintervention: Nonsignificant weak correlation between surgical technical error and team NOTECHS scores Significant moderate negative correlation between team SA and surgical technical errors Significant strong negative correlation between surgical SA and surgical technical error Postintervention: Significant weak negative correlation between surgical technical error and surgical and team NOTECHS scores; significant stronger negative correlation between surgical technical errors and surgical NOTECHS scores</td>
<td>Concurrent use of NOTECHS and OCHRA by same observer may lead to greater agreement between scales. Single site study Single general surgical procedure Small sample size</td>
</tr>
</tbody>
</table>

LM, leadership and management; NOPEs, Non-Operative Procedural Errors; OCHRA, Observational Clinical Human Reliability Analysis; OR, operating room; Oxford NOTECHS, Oxford Non-Technical Skills; PD, problem solving and decision making; SA, situational awareness; TC, teamwork and cooperation.

*Italics denote author critical appraisal.
scoring leadership and management in OR nurses were associated with fewer procedural problems and errors (eg, dropping a sterile piece of specialist equipment, forgetting to connect equipment power leads, administering the wrong drug), higher levels of these same skills among surgeons decreased overall operative time. Two additional studies recorded operative events via real-time observation and subsequently classified them as surgical errors and teamwork failures or disruptions of various types. A strong positive correlation \( (r = 0.67, p < 0.001) \) between technical error and teamwork failure was found – the latter defined as operative events constituting teamwork failures according to human factors theory.

The impact of nontechnical skills on technical performance of anesthesiologists and OR nurses

Only 1 of the 28 studies in this review (not tabulated) assessed the impact of nontechnical skills on anesthesiologists’ technical performance, and found a strong positive correlation between the 2 \( (r = 0.73, p < 0.001) \). This was an intervention study, conducted in a simulated environment. Forty-two anesthesiologists’ nontechnical skills were assessed using the validated Anesthetists Non-Technical Skills (ANTS) tool, which measures resource management, planning, leadership, and communication. Technical performance was evaluated using the Medical Management Checklist. This study also reported only overall correlations, without skill-specific analyses. Our search did not identify any studies that assessed the impact of nurses’ nontechnical skills on their own technical performance in the OR.

DISCUSSION

This is the first systematic review on the impact of nontechnical skills on technical performance in surgery. A striking feature of the studies included in this review was the wide range of nontechnical aspects of performance assessed. Dealing effectively with stress and fatigue, and seeking performance feedback have traditionally been viewed as performance management skills, but have more recently been identified as core nontechnical skills, and they reach far beyond the behavioral and cognitive elements, typically associated with the term nontechnical skills, such as communication and team working. Despite the significant heterogeneity in the evidence base, the following findings emerged regarding how surgeons’ nontechnical skills may influence their technical performance:

1. There is no evidence that poor communication in the OR negatively affects technical performance. The fact that communication is not explicitly included in all nontechnical assessment tools may contribute to the lack of relationship between the 2 skills found in this review.
2. Failures in nontechnical skills (especially in situational awareness among surgeons) are associated with a higher rate of technical errors.
3. Coping with the deleterious impact of excessive levels of stress in the OR is key to maintaining optimum technical proficiency. This finding complements a recent systematic review that highlighted the significant impact of stress on surgical performance. The impact of stress depends on the level of expertise of the surgeon and the nature of the task.
4. Increased levels of fatigue are associated with detriments to particular aspects of technical performance. The effect, however, appears to be task dependent.

Although we also sought to investigate the impact of other team members’ nontechnical skills on technical outcomes of surgery, this was limited by the available literature, in which evidence on other team members in the OR, aside from surgeons, was conspicuously absent. This is a serious concern because safe surgical performance can be dependent on the successful interaction of different people working together in the OR. This is especially true in the case of crises, when the performance of a surgeon can be significantly enhanced or impeded by the team skills of anesthetic and nursing personnel. Furthermore, focusing on the surgeons’ nontechnical skills alone, to the exclusion of others, is not a true representation of teamwork. So it remains to be empirically investigated how interprofessional team interactions affect surgical performance and subsequently, how they can be optimized through team training to improve patient outcomes.

With regard to the secondary aim of this review, vast variability in assessment tools was noted. Assessment of technical performance ranged from using tools with extensive evidence of validity and reliability, such as the Objective Structured Assessment of Technical Skills (OSATS) to rather crude measures of technical performance, such as time taken to complete a task (speed does not always equal accuracy or safety). The same was true for nontechnical assessment, which ranged from tools with extensive evidence of validity and reliability, such as the Observational Teamwork Assessment for Surgery (OTAS) to study-specific scales with no validity evidence to support them. Use of tools with evidence of validity and reliability is a prerequisite for successful scientific measurement of a skill or of a performance parameter in question. Poorly validated metrics present a significant risk of missing a true correlation between technical and nontechnical perfor-
formance because of a lack of sensitivity or a lack of reliability in the metrics (ie, a type I error). Better validated tools capture underlying skill or performance more accurately, minimizing error and bias in data collection and increasing confidence in the validity of the findings. With the current availability of assessment tools that are psychometrically robust to capture nontechnical skills, and with recent systematic reviews detailing the psychometric robustness of technical and nontechnical assessment measures, we strongly recommend that future research consistently use such tools to improve the robustness of the evidence base. Moreover, tool use by trained assessors should be considered a prerequisite for future research. Training assessors is part of establishing validity of the assessment and quality control; it ensures assessments are truly comparable within and between studies. This is particularly true for nontechnical assessments because behaviors are harder to assess than technical performance. This need has been identified in the surgical literature, and valid training packages for assessors are published to address the gap.

Limitations
In the light of the evidence base, certain limitations become apparent. A key limitation is that although the relationship between nontechnical skills and technical performance has attracted much attention in the surgical literature, many studies have failed to address this in a consistent manner. Twenty articles were excluded from this review because even though they did report descriptive assessments of both nontechnical skills and technical performance, they did not report any associations between the 2 skills sets using statistical analyses. This is an important finding in itself, for 2 reasons. First, it may highlight publication bias if correlational analyses were not reported due to nonsignificant findings. Second, it highlights significant variability in reporting across studies, rendering cross-study comparisons difficult. Therefore, with a view toward improving the standard of the evidence base, we advocate that both descriptive and correlational analyses be provided for any study jointly assessing technical and nontechnical performance. A further limitation of this review is that researchers typically treat all nontechnical skills as a single entity, so they report only single, global ratings of these skills. The failure to report findings at the level of individual nontechnical skills prevents firm conclusions from being drawn regarding the contribution of each particular skill to technical outcomes. Future studies should report assessments of individual skills, as well as global scores. This will allow the relative importance of each individual nontechnical skill to be elucidated, leading to targeted training and improvement efforts. Furthermore, the majority of the studies in this review were conducted in simulated environments; the validity of these studies clearly rests on the assumption that performance in a simulated environment is directly comparable to performance in the OR. In addition, the studies that were conducted in the OR concentrated on routine elective procedures; there may be fundamental differences in the impact of nontechnical skills on particular aspects of technical performance, in different procedures, and also in emergency surgery settings. At present, this is unclear and future research should empirically investigate such gaps in the literature. These limitations notwithstanding, this is the first systematic review that synthesizes the current evidence base regarding the impact of nontechnical skills on surgical performance.

Implications and recommendations for surgical practice and training
Overall, this review provides evidence that the nontechnical skills of OR teams do have an effect on their technical performance. Taken together with other recent reviews, our review suggests that training to improve nontechnical skills has the potential to improve team work, performance, and safety in the OR, and therefore to ultimately contribute positively to patient outcomes. As our understanding of the interactions between specific skills and technical performance increases, tailor-made training packages aiming at specific skills for specific grades of surgical expertise can be developed, implemented, and evaluated. Within the current training paradigm, incorporating nontechnical elements into the feedback provided to trainees (ie, in addition to technical feedback) could be a first step toward raising awareness and facilitating behavior change in the workplace.

Implications and recommendations for future research on nontechnical skills within surgery
Future research should address the issue of variability in assessment tools. Recent reviews have compared assessment tools for both technical and nontechnical performance in relation to scale formulation, validity, reliability, and feasibility. These should be used to inform future tool selection for further studies rather than ad-hoc development on a study-by-study basis. Reporting of findings should also be more consistent, covering associations between different skill sets.

Taking a wider perspective, interventions designed to improve nontechnical skills in the OR have started to emerge. In the UK, the Surgical Safety Checklist that emerged from an international World Health Organization pilot study has been mandated for use in all ORs since January 2009. The key aim of the checklist is to enhance team communication and coordination, and
therefore improve patient safety. In the US, several modules are included in Phase III of the American College of Surgeons and the Association of Program Directors in Surgery (ACS-APDS) Surgical Skills Curriculum13 and several of the Accreditation Council for Graduate Medical Education (ACGME) competencies.60 These developments are very encouraging. Coupled with scientifically sound evaluations of their impact on team and technical performance using validated metrics, they are likely to lead to a new generation of OR teams with technically proficient and effective team members.

CONCLUSIONS
In the past 5 years, nontechnical skills have become a prominent feature of the surgical literature and also, to a lesser extent, of the surgical curricula. Although the available literature is somewhat heterogeneous, this review provides evidence that these skills can and do have an effect on surgeons’ technical performance. Future research should approach the assessment of nontechnical skills in a more standardized and scientific manner. Providing evidence-based training in nontechnical skills, alongside psychometric robust tools for their measurement can lead to a new generation of surgeons competent in all the skills required for safe, high quality patient care.

Appendix. Search Strategy
Category A = ‘Surg$’ OR ‘nurs$’ OR ‘anaesthes$’ OR ‘anesthes$’ OR ‘operating theatre’ OR ‘physician$’ (MeSH) OR ‘physician assistant$’ (MeSH) OR ‘nursing staff$’ (MeSH) OR ‘nurse$’ (MeSH) OR ‘nurse anesthetist$’ (MeSH) OR ‘nurse administrator$’ (MeSH) OR ‘nurse clinician$’ (MeSH) OR ‘nurse$’ (MeSH) OR ‘operating department practitioner$’ OR ‘operating department assistant$’
Category B = “technical performance” OR “technical skills$” OR “dexterity” OR “psychomotor performance$” (MeSH) OR “motor skills$” (MeSH) OR “motor performance” OR “technical error”
Category C = ‘non-technical performance’ OR ‘non-technical skills’ OR ‘interpersonal skills$’ OR ‘communication’ OR ‘leadership’ OR ‘teamwork’ OR ‘briefing’ OR ‘planning’ OR ‘preparation’ OR ‘res’ OR ‘advice’ OR ‘feedback’ OR ‘stress’ OR ‘pressure’ OR ‘fatigue’ OR ‘cognitive skills$’ OR ‘situational awareness’ OR ‘mental readiness’ OR ‘assessing risks$’ OR ‘anticipating problems’ OR ‘decision making’ (MeSH) OR ‘adaptive strategies’ OR ‘adaptive flexibility’ OR ‘workload distribution’

Limit: Publication date: 1980-Week 2, April 2010, English, Humans

Author Contributions
Study conception and design: Hull, Arora, Aggarwal, Sevdalis
Acquisition of data: Hull, Arora
Analysis and interpretation of data: Hull, Arora, Sevdalis
Drafting of manuscript: Hull, Arora, Sevdalis

Critical revision: Vincent, Darzi, Aggarwal

REFERENCES


