A systematic review of the methodological quality and outcomes of RCTs to teach medical undergraduates surgical and emergency procedures

Roger E. Thomas, MD, PhD;* Rodney Crutcher, MD;* Diane Lorenzetti, MLS†

Background: There is no systematic review of the methodological quality of randomized controlled trials (RCTs) of teaching surgical and emergency skills to undergraduates. **Methods:** We searched the Cochrane Collaboration Controlled Trials Register, the Cochrane Database of Systematic Reviews, MEDLINE, EMBASE, ERIC, DARE and the University of Toronto Continuing Medical Education database for RCTs in all languages. **Results:** We identified 19 RCTs. Four tested methods of IV access, 1 found intraosseous access faster than the umbilical vein in neonates, and 1 found that one type of intraosseous needle had higher success rates. Two RCTs of intubation skills did not identify a superior technique. One RCT of CPR found video instruction superior to the American Heart Association Heartsaver course. Of 2 RCTs of trauma skills, 1 found no improvement and 1 found improvement only on the day of instruction. One RCT found both computer and seminar training improved epistaxis management. One RCT gave students preoperative anatomy instruction, and they received higher ratings from surgeons. One RCT asked students to study surgical scenarios preoperatively, and they improved their surgical intensive care unit skills. One RCT gave students video and paper-cut instruction of the Whipple procedure; both groups improved, but there were no differences between groups. One RCT taught uteteroscopy and stone extraction and found groups that used low- and high-fidelity bench models improved, compared with the didactic group. Four of 5 RCTs of knot tying showed improvement. **Conclusions:** This systematic review assessed the quality of RCTs used in teaching undergraduates surgical and emergency skills. There are many positive study outcomes, but there are significant methodological weaknesses in the study design. Students varied in their skills, and most did not demonstrate optimal performance in any of the procedures. This review provides a baseline for further work important to both medical education and clinical practice.

Contexte : Il n’existe pas d’analyse systématique de la qualité méthodologique des études contrôlées randomisées (ECR) de l’enseignement des techniques de chirurgie et d’intervention d’urgence au premier cycle. **Méthodes :** Nous avons cherché des ECR dans toutes les langues dans le registre Cochrane Collaboration Controlled Trials Register, dans la base de données Cochrane Database of Systematic Reviews, dans MEDLINE, EMBASE, ERIC et DARE, et dans la base de données sur l’éducation médicale continue de l’Université de Toronto. **Résultats :** Nous avons trouvé 19 ECR. Quatre portaient sur des méthodes d’accès IV. Dans un cas, on a découvert un accès intraosseux plus rapide que la veine ombilicale chez les nouveau-nés et dans un autre, on a constaté qu’un type d’aiguille intraosseuse produisait de meilleurs taux de réussite. Deux ECR portant sur les techniques d’intubation n’ont pas dégagé de technique meilleure. Une ECR portant sur la RCR a révélé que la formation vidéo donnait un meilleur résultat que le cours Heartsaver de l’American Heart Association. Deux ECR portaient sur les techniques de traumatologie : on n’a constaté aucune amélioration dans un cas et dans l’autre, on a constaté une amélioration le jour de la formation seulement. Une ECR a révélé que la formation par ordinateur et par colloque améliorait la prise en charge de l’épistaxis. Une ECR a donné aux étudiants une formation en anatomie préopératoire et ils ont reçu des notes plus élevées des chirurgiens. Dans le cadre d’une ECR, on a demandé aux étudiants d’analyser des scénarios chirurgicaux avant l’intervention et ils ont amélioré

*From the *Department of Family Medicine and the †Institute of Health Economics and Centre for Health and Policy Studies, Department of Community Health Sciences, University of Calgary, Calgary, Alta.*

Accepted for publication Oct. 21, 2005

Correspondence to: *Dr. Roger Thomas, Department of Family Medicine, University of Calgary, University of Calgary Medical Clinics, 1709–1632 14th Ave. N.W., Calgary AB T2N 1M7; fax 403 210-9204; rthomas@ucalgary.ca*

leur technique chirurgicale aux soins intensifs. Au cours d'une ECR, on a donné aux étudiants des instructions par vidéo et sur papier au sujet de l'intervention Whipple : les deux groupes se sont améliorés, mais on n'a pas constaté de différence entre les deux. Au cours d'une ECR, on a enseigné l'urétéroscopie et l'extraction de calculs et constaté que les groupes qui utilisaient des mannequins basse et haute fidélité s'amélioraient comparativement au groupe qui recevait de l'enseignement didactique. Quatre des cinq ECR portant sur le nouage ont révélé une amélioration. **Conclusions :** Cette analyse systématique a évalué la qualité des ECR utilisés dans l'enseignement des techniques de chirurgie et d'urgence au niveau du premier cycle. Les résultats positifs sont nombreux, mais la conception des études présente d'importantes faiblesses méthodologiques. L'habileté des étudiants variait et la plupart n'ont pas produit un rendement optimal, quelle que soit l'intervention. Cette analyse constitue un niveau de référence pour d'autres travaux importants à la fois pour la formation en médecine et pour la pratique clinique.

There are several randomized controlled trials (RCTs) of teaching undergraduates fundamental surgical and emergency skills that intended to assess whether the teaching of these skills can be improved, but there is no systematic review that evaluates the quality and findings of this teaching. We use the international Quality of Reporting of Meta-Analyses (QUOROM)[1] statement to assess the methodological quality of these RCTs, possible sources of bias and whether the conclusions drawn by the authors can be relied on by surgical teachers to improve the teaching of undergraduates.

Methods

Literature search

For RCTs in all languages, we searched the Cochrane Controlled Trials Register, using the term "medical student." We searched MEDLINE using the following terms: "medical students" and "randomized controlled trials" or "systematic reviews," "meta-analysis," "crossover studies," "intervention studies," "Latin squares," "factorial," "multicentre studies," "cohort studies," "prospective studies" or "longitudinal studies" (and spelling variations of these terms). We used similar terms to search DARE, EMBASE, the University of Toronto Continuing Medical Education database and ERIC.

Study selection

Two reviewers independently assessed whether the study was an RCT and taught surgical or emergency procedures to medical students. We excluded any study where the outcomes for medical students could not be separated from other health professional groups, or where a common fundamental surgical or emergency procedure was not taught. Surgical procedure simulators and virtual reality simulators are generally not used to teach fundamental procedures to undergraduates worldwide, thus we excluded RCTs of simulators.

Validity assessment

All studies that appeared from their titles or abstracts to be RCTs, or where the abstract did not reveal a decision about the study design, were evaluated by independent assessment of the full text of each study.

RCTs were categorized according to the criteria of the Cochrane Collaboration Reviewers' Handbook[2] as having low, moderate or high risk of bias according to methodological strength. We based our estimate of bias on the 4 Cochrane criteria for minimizing bias:

1) Selection bias. We assessed the study as being at low risk of bias if participants were randomly assigned to experimental or control groups. We assessed whether randomization was concealed from the experimenters.
2) Performance bias (inadequate delivery of the intervention). We noted whether a process analysis was performed, to assess whether the interventions were fully delivered to all participants according to the study protocol. We also assessed whether membership in the intervention or control groups was blinded to the participants and experimenters.
3) Attrition bias. If an analysis was not performed or if known biasing effects of attrition were not adjusted for in the analysis, attrition bias was considered likely. In this case, the study was considered to be at moderate risk of bias.
4) Ascertainment bias (if studies did not use the same methods of ascertainment for both experimental and control groups). We also assessed whether ascertainment of outcomes in the intervention or control groups was blinded to the experimenters.

We assessed 3 additional aspects of study design that affect the quality of RCTs and that are also common problems in the field of medical education studies:

5) Inadequate sample size. If the results for the key hypotheses were statistically significant, the study was assessed as at low risk of bias for type II error, even if the study did not have a power computation. If the results were negative and there was no power computation, we assessed the study as at risk of type II error.
6) Intention-to-treat analysis. If the authors did not plan an intention-to-treat analysis and there was no attrition analysis showing that loss of subjects from the experimental and control groups did not affect the outcomes, we assessed the study as being at risk of overestimating the effects of interventions.
7) Statistical bias. Studies that ran-

domize by cluster (group) but analyze at the level of the individual are at risk of drawing false positive conclusions because part of the outcome may be due to discussions between class members. The cluster is now the sampling unit and not the individual. Failure to take account of clustering and the size of interclass correlations may lead to inadequate sample size and the risk of drawing false nonsignificant conclusions (type II error).[3–6] We assessed studies as at moderate risk of bias if they did not control for clustering.

The Cochrane Reviewers' Handbook[2] recommends the following approach to summarizing the risk of bias:

Assessment of the risk of bias in RCTs:

A. Low risk – plausible bias unlikely to seriously alter the results; all of the criteria met.
B. Moderate risk – plausible bias that raises some doubt about the results; 1 or more criteria partly met.
C. High risk – plausible bias that seriously weakens confidence in the results; 1 or more criteria not met.

The Handbook states the following:

> The relationships suggested above will most likely be appropriate if only a few assessment criteria are used and if all the criteria address only substantive, important threats to the validity of study results.[2]

Based on our assessment that these 4 sources of bias and 3 additional aspects of study quality might threaten the validity of a study, studies were assigned to 3 categories: low risk, moderate risk and high risk of bias. In synthesizing the results, conclusions were based on those with low or moderate risk of bias.

Data were independently extracted by the 2 reviewers, and discussion continued until agreement was achieved.

Based on considerable heterogeneity in study design, intervention, outcome measures and statistical reporting, we determined that quantitative synthesis was not appropriate, and we used a narrative systematic review.

Results

Trial flow

We identified 88 potential RCTs. On examination of the full text, 21 were excluded because they were not RCTs[7–27] and 48 because they were either not on the topic of learning procedures in surgery and emergency medicine, or the outcomes for medical students could not be identified.[28–75] Nineteen RCTs remained for evaluation (Table 1).[76–94] No replications of RCTs were identified (see Fig. 1 for QUOROM flow chart).

Methodological quality

Strengths

The strengths of these 19 RCTs are as follows: a) complete delivery of the experimental stimuli, because the interventions were closely supervised by faculty; b) minimal drop-outs, because the students completed the experiments as part of required university courses. (Although no study planned an intention-to-treat analysis, most accomplished one because of these high completion rates.); and c) although most studies did not blind students, researchers or assessors, most of the outcomes were performance times, success in a procedure, or scores on a multiple choice ques-

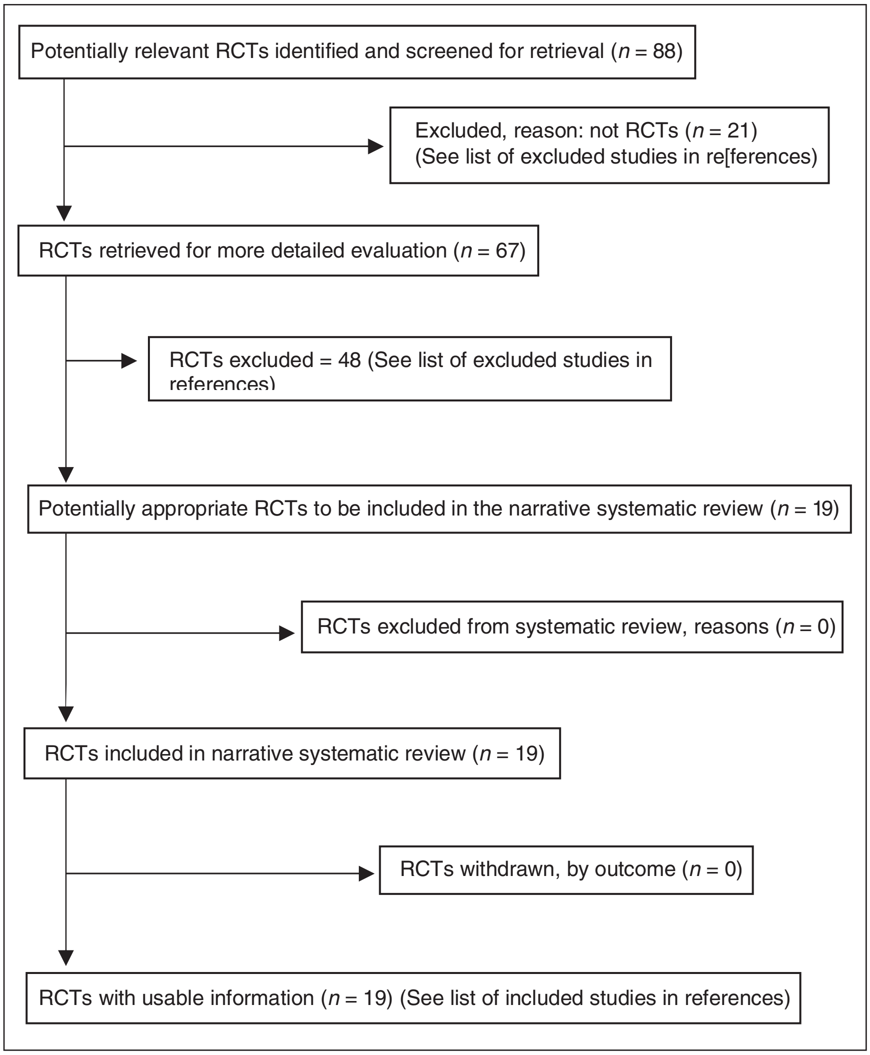

FIG. 1. QUOROM flow chart.

tionnaire (MCQ); thus the absence of blinding might have introduced little bias, except where surgeons assessed the quality of work using personal judgement and were unblinded Table 2 summarizes each study's methodological quality.

Weaknesses

The weaknesses of the RCTs are as follows: a) although all 19 were described in the text as RCTs, only 5 described the actual method of randomization (Abe et al[76] used coin toss, Ali et al[77] alternated removing names from a box, the students in the study by Carr et al[78] chose from face-down cards, Matsumoto et al[79] chose by candidate number and Todd et al[80] chose from a table of random numbers). Only Abe and colleagues[76] and Todd and colleagues[80] used a strong method of randomization. Only From and colleagues[81] and Todd and colleagues[80] concealed randomization and only Matsumoto and colleagues[79] reported a power computation, and Carr and colleagues reported a post-hoc computation.[78] None of the studies blinded participants; 4 blinded instructors (From et al,[81] Matsumoto et al,[79] Rogers et al,[95] Todd et al[80]), and 4 blinded assessors (From et al,[81] Hong et al,[82] Matsumoto et al,[79] Todd et al[80]). None of the studies described possible cointerventions during the study period (these would have been improbable during brief experiments but likely during rotations lasting up to 12 wk). If students were assigned to groups, it was difficult from the descriptions to assess how much communication between students occurred during the interventions and, thus, how much of the result was due to learning from fellow students, in addition to the program. Based on the description of how groups operated, it appears that 7 did not allow for the effects of clustering in the statistical analysis (Carr et al,[78] From et al,[81] Gilbart et al,[83] Rogers et al,[84,85] Summers et al,[86] Todd et al[80]). In assessing whether these RCTs are subject to bias, failure to deliver the intervention would be minimal because of the close observation and certainty that the interventions were delivered to all participants; bias due to attrition would also be minimal because nearly all participants completed the experiments. Weight should be given to the low incidence of blinding and concealment, the absence of power computations (inadequate sample size could be the reason for nonsignificant results) and failure to adjust for clustering in the statistical analysis (Table 3).

Table 1

RCTs of teaching topics in surgery and emergency medicine to medical students

Topic	No. of RCTs
Airway management	2
Basic surgical techniques	10
CPR	1
IV access	4
Trauma assessment	2
Total	**19**

RCT = randomized controlled trial; CPR = cardiopulmonary resuscitation; IV = intravenous.

Previous research

Four authors state that they identified no previous studies to guide their research (From et al,[81] Hong et al,[82] Matsumoto et al,[79] Talan et al[88]); 5 studies conducted literature searches and analyzed the studies to guide their own research design (Carr et al,[78] Rogers et al,[84,93,95] Summers et al[86]; 9 authors cited studies but did not build on them to improve the design or execution of their studies or only cited them in the concluding discussion section (Abe et al,[76] Ali et al,[77] Gilbart et al,[83] Jun et al,[89] Mann et al,[91] Petroianu et al,[90] Rogers et al,[85] Rogers et al,[94] Todd et al[80]). One author (Engum et al[87]) cited none of the relevant previous RCTs identified in our review. None of the studies had a section titled "literature search," none stated which databases were searched or the search terms used, and none mentioned whether they consulted a health sciences librarian or expert to identify studies.

Although medical students vary in their skills and application, no author described difficulties or ease in instruction, thus we can learn nothing from these studies about which aspects of the procedures were more difficult for some students, how to help students who encounter problems or whether the instruction technique for these procedures requires modification or improvement to help specific students.

Outcome measures

Most of the outcome measures had face validity (Table 2). When authors used scales, few described their scales in sufficient detail or printed them out in the text so that readers could understand each step in the learning process.

As outcome measures, Abe and colleagues,[76] Jun and colleagues[89] and Talan and colleagues[88] used time taken and success in identifying the vein, and Engum and colleagues[87] described their success in accessing the vein. Petroianu and others[90] measured time to intubation and success, and Todd and others[80] assessed intubation on a 5-point scale from 1 (not competent) to 5 (outstanding).

Researchers who used objective structured clinical exam (OSCE) stations used OSCE scores; Gilbart and colleagues,[83] Rogers and colleagues[94] and Ali and colleagues[77] used trauma evaluation and management (TEAM) protocol scores to assess the adequacy of advanced trauma life support (ATLS) learning; Hong and colleagues[82] used surgeons' performance ratings, and Mann and colleagues[91] used scores of understanding the anatomic steps in the Whipple procedure. Rogers and colleagues[95] evaluated students' knot tying by independent evaluation of videotapes by

Table 2

Summary of the RCT study design, outcome measures, subjects and results

RCT	Country	Study design	No. of subjects: Baseline	No. of subjects: Study end	Outcome measures	Results
IV access						
Abe et al[76]	USA	1) Randomized to intraosseous access either in a turkey leg or plastic model of an infant's leg 2) Comparison with access to neonatal umbilical vein Unlimited practice permitted	42 medical students in Hawaii	42	Time to successful placement of an IV line VAS of difficulty (0–10)	Initial attempt without experience: no difference between intraosseous methods, combined intraosseous 52 s v. umbilical vein 134 s ($p < 0.001$) (success rate 95% v. 79% (ns)) Experienced attempts: no difference between intraosseous techniques, combined intraosseous 45 s v. umbilicus 95 s ($p < 0.001$); success rate 93% v. 88% (ns) VAS of difficulty significantly lower for intraosseous (level not stated)
Engum et al[87]	USA	Randomized to computer IV simulator compared with IV placement in volunteer's arm	93, 3rd-yr medical students in Indiana (47% received previous IV instruction)	93	21-item checklist	Traditional group score 14.02; computer-trained group 12.56 for the 93 medical students + 70 nursing students scores were significantly higher with the traditional method ($p = 0.01$)
Jun et al[89]	USA	Randomized to use standard bone marrow needle or screw-tipped intra-osseous needle in turkey femur or pork ribs	42 medical students in Hawaii	42	Time to successful catherization VAS of difficulty (0–10)	For inexperienced attempt, time with standard bone marrow needle was 33 s and for the screw-tipped needle 54 s ($p < 0.019$); success rates of 83% and 76% (ns) After training, 32 v. 27 s (ns) and success rates 79% v. 95% ($p < 0.05$)
Talan et al[88]	USA	Randomized to cutdown of cadaver cephalic vein at wrist or saphenous cut-down at ankle then crossover	17 medical students in Los Angeles	17	Time to isolation of vein	Average time to isolate cephalic vein 85 s v. saphenous vein 70 s (ns) No nerve artery or tendon injuries with either technique
Airway management						
From et al[81]	USA	Groups of 5–8 3rd-yr students randomized to American Heart Association self-study course with sensorised mannequin compared with lectured with guided practice on standard tracheal intubating mannequin	97, 3rd-yr medical students in Iowa	97	Airway management skills rated 4 (excellent) to 1 (inadequate)	No statistically significant differences on overall assessment of mask airway efforts (didactic 3.36, learning system 3.34); intubation efforts (3.15, 3.10) or mask and intubation (3.25, 3.27)
Petroianu et al[90]	Germany	Randomization to 5 nasal or 5 oral intubations of Laerdal Airway Management; trainer using a Trachlight	24 medical students, Univ. of Heidelberg	24	Time to intubation; successful intubation as measured by lung inflation	No differences in time to intubate or success on 10th attempt nasally or orally

Continued …

Table 2 (continued)

RCT	Country	Study design	No. of subjects: Baseline	No. of subjects: Study end	Outcome measures	Results
Todd et al[80]	USA	Comparison of 34-min video self-instruction for CPR with 4-h American Heart Association Citizen Heartsaver A course; groups of 16	91 medical students in Atlanta	89*	CPR on average 100 d after training in a simulated setting rated on an ordinal scale from 1 (not competent) to 5 (outstanding); each skill step was rated adequate or not	20 of 47 (43%) trainees who took the Heartsaver course were judged not competent, 8 of 42 (19%) who took the VSI course (risk difference 24%, 95% CI 5%–42%) were judged not competent Median scores were 2 (questionable competence) for Heartsaver and 3 (competent) for VSI ($p = 0.04$) For 11 of 14 skill steps VSI achieved the same or superior scores to Heartsaver VSI accomplished more correct ventilations (3.6) than Heartsaver (1.9; $p = 0.004$)
Trauma assessment						
Ali et al[77]	Australia	Randomization of volunteers to 4- to 6-h TEAM trauma evaluation and management program or control	73, final-yr medical students, Melbourne	73†	MCQs from ATLS question bank	Statistically significant improvement in the scores of the experimental compared with the control groups ($p < 0.05$) during the day of instruction
Gilbart et al[83]	Canada	Randomized to groups of 6 for 2-h computer-based CAE trauma simulator to perform resuscitation steps or 2-h seminars to verbalize resuscitation steps. 2 nonrandomized control groups received no additional teaching‡	139 (79%) of 176, 4th-yr clinical clerks, Univ. of Toronto	107 (61%); 57 simulator, 50 seminar; 32 C1; 37 C0	Scores on 2 trauma and 5 nontrauma OSCE stations; scores on post-encounter probes when students were asked questions	All nontrauma stations: seminar 7.53, simulator 7.46 (ns) All trauma stations: seminar 7.78, simulator 7.93 (ns) On post-examination probe: all students performed better on the 2 trauma than the 5 nontrauma stations ($p < 0.05$); no differences between grps Both teaching groups performed better on the 2 trauma than the 5 nontrauma stations ($p < 0.05$), with no significant differences between the 2 types of teaching There were no significant improvements in scores for stations where transfer of skills could be expected
Surgical technique						
Carr et al[78]	USA	Randomization to a pretest, computer module or small group seminar to learn management of epistaxis	58 medical students, Univ. of Toronto	58	17-item short answer test; anterior nasal packing on model, assessed by 16-item checklist	Pretest written 28%, practical 22% Computer written 84%, practical 93% Seminar written 89%, practical 87% Computer and seminar groups had higher scores than the pretest group ($p < 0.05$)
Hong et al[82]	Canada	4 students received 10–15 min oral presentations, used interactive anatomy software on a computer about the anatomy relevant to the operation and attended 2 operations; 4 received no instruction; the groups were then crossed over	8, 4th-yr medical students during the 8-wk surgical rotation at the Toronto Hospital 1994–1995	8 students assigned to 2 operations	Senior surgeon's assessment of the clerk's knowledge and understanding of the operation (6 items) and the clerk's assessment of the experience in the operating room (8 items)	The surgeons' ratings were higher for students who received the preoperative tutorial (3.7 v. 3.0; $p < 0.01$). The clerks rated their experience more positively if they received the preoperative tutorial (4.0 v. 3.1; $p < 0.001$)

Continued …

Table 2 (continued)

RCT	Country	Study design	No. of subjects: Baseline	No. of subjects: Study end	Outcome measures	Results
Mann et al[91]	USA	Randomized to video or paper-cut exercise to learn the Whipple procedure and posttest; groups switched exercises then took a second posttest	37 medical students in Philadelphia	37	Scores on 1 pretest and 2 posttests	No difference between groups' pre-test scores (p = 0.290). After the first exercise, students who performed the paper-cut showed greater improvement in test scores compared with students who saw the video (p = 0.0035). After both groups had completed exercises, mean changes from baseline were no longer different (p = 0.58)
Matsumoto et al[79]	Canada	Randomized to learn ureteroscopy and extraction of a stone from the midureter by a 1-h didactic teaching by urologist or 1-h sessions supervised by a urologist of low- or high-fidelity bench model practice	40 medical students in Toronto	40	Global rating scale; checklist; pass rating; time to complete the task	Significant effect of hands-on training (p < 0.01); low-fidelity group had significantly higher global rating scores (p = 0.012), checklist scores (p = 0.006), pass ratings (p = 0.001) and shorter time to complete the task (p = 013) than the didactic group No significant differences between the high- and low-fidelity groups
Rogers et al[95]	USA	Study I: medical students randomized to learn how to tie knots to 1 of 4 training groups (no errors, errors only, correct only, errors and correct) Study II: students rated 24 taped examples of knot tying	30 medical students, MCG	30†	Study I: students were videotaped before and after instruction and rated blindly by 8 surgeons on a 7-point rating scale Study II: students rated 24 taped examples of knot tying; 8 surgical faculty randomized to receive training in reviewing errors or not	The 4 common errors: too much motion in right hand (38%), failure to maintain constant tension (17%), hands too close to knot (13%), failure to cross hands (7%) Study 1: significant improvement only in the "error plus correct" training group (p < 0.01) Study 2: student IRR was lower for the "correct only" group (p < 0.01) compared with the other 3 groups; no differences between faculty in IRR
Rogers et al[93]	USA	Randomized to 1-h computer-assisted teaching how to tie a square knot or computer assisted + individualized expert feedback	108 medical students, Univ. of Georgia	105	Best and fastest pair of 2-handed square knots; each taped session independently reviewed by 3 surgical faculty on a scale (max 24)	For the computer group pretest scores 2.6, post 12.0 (p < 0.001) For computer + feedback pretest 3.1; posttest 15.8 (p < 0.001) Computer + feedback improved more than computer (p < 0.001)
Rogers et al[84]	USA	Randomized to learn how to tie surgical knots either in groups of 6–8 with each student using an individual computer for assisted instruction or in pairs, with each pair using 1 computer	77 medical students, MCG	77	Best and fastest pair of 2-handed square knots before and after instruction, assessed independently by 3 surgical faculty with a checklist (max score 24)	The proportion of knots squared increased in the computer group: pretest 11%, posttest 60% (p < 0.001) global rating 3.95; in the computer + feedback by peer group: pretest 10%, posttest 81% (p < 0.001), global rating 4.08 Computer group had a significantly lower % of knots squared (p = 0.04) No significant differences between groups in global ratings
Rogers et al[85]	USA	Randomized to either 1-h computer instruction or lecture with slides of knot tying and feedback session to learn how to tie surgical knots	91 medical students, MCG; 8 excluded for operating or ED experience, 1 as the video was not useful	82	Best and fastest pairs of 2-handed square knots before and after instruction were blindly and independently evaluated by 3 surgical faculty members with a checklist (max score 24) with Cronbach α = 0.79	The proportion of correctly tied square knots was 85% in the computer group and 90% in the lecture group (ns) Average time was 19.6 s in the computer and 17.4 s in the lecture group (ns) Performance score was 12.8 in the computer and 17.4 in the lecture group p < 0.001)

Continued …

Table 2 (continued)

RCT	Country	Study design	No. of subjects: Baseline	No. of subjects: Study end	Outcome measures	Results
Rogers et al[94]	USA	Randomized to 1 of 2 clinical scenarios to learn critical care medicine skills before elective, then crossed over to other scenario after the elective	40 medical students on the critical care medicine elective in 5 surgical intensive care units in Pittsburgh	40	OSCE with checklist of key behaviours for 5 stations: 1) airway, breathing and circulation 2) prepare a mannequin for intubation, obtain an acceptable airway, demonstrate bag-mouth resuscitation, perform acceptable laryngoscopy and intubation 3) provide appropriate mechanical ventilator settings 4) manage hypo-tension 5) request and interpret pulmonary artery data and initiate appropriate therapy	The mean pre-elective score was 59.0% (SD 8.3%), compared with 85.9% (SD 7.4%) ($p < 0.0001$) after the elective Scores for the OSCE stations were: 1) pre 9.6, post 9.8 (max 12) 2) pre 13.7, post 20.4; $p < 0.01$ (max 24) 3) pre 3.9, post 8.8 (max 9); 4) pre 3.8, post 5.2; $p < 0.01$ (max 9)
Summers et al[86]	USA	Randomized to didactic surgical skills manual, videotape or multimedia computer training for basic surgical skills (5 knot-tying and 7 suturing skills)	69 medical students, Medical College of Wisconsin	69 after instruction, 58 1 mo later	After the instructional session and 1 mo later, each student performed each technique on fresh pigs' feet; each technique was rated on instrument handling, body position, accuracy, tightness, alignment and time to perform Two assessors were randomly assigned to rate half of each group Students were allowed 1 min for each knot-tying and 2 min for each suturing task	1) Immediate follow-up: a) % correct: didactic 78%, video 80%, computer 84% (didactic v. compute;r $p < 0.01$) b) % complete: didactic 92%, video 94%, computer 97% (didactic v. computer; $p < 0.01$) c) performance quotient score (ns) d) MCQ exam score: didactic 63%, videotape 49%, computer 49% (didactic v. computer; $p < 0.01$). 2) Follow-up at 1 mo: a) % correct: didactic 88%, video 90%, computer 89% (ns) b) % complete: didactic 89%, video 94%, computer 93% (didactic v. video; $p < 0.01$) c) performance quotient: didactic 396, video 413, computer 427 (didactic v. computer; $p < 0.01$). d) MCQ exam: didactic group higher scores than video or computer ($p < 0.01$).

RCT = randomized controlled trial; VAS = visual analogue score; ns = nonsignificant; IV = intravenous; CPR = cardiopulmonary resuscitation; CI = confidence interval; VSI = video self-instruction; TEAM = trauma management and evaluation; MCQ = multiple choice questionnaire; ATLS = advanced trauma life support; CAE = CAE electronics patient simulator system; C1 = comparison group 1; C0 = comparison group 0; OSCE = objective structured clinical examination; MCG = Medical College of Georgia; IRR = interrater reliability; ED = emergency department; SD = standard deviation.
*2 did not complete testing.
†Text did not state any students dropped out.
‡Comparison 1 = students who volunteered but were not able to participate in either of the above courses; comparison 2 = students who did not volunteer.

3 surgeons, using a 7-point rating scale. Three other studies[93,84,85] used independent evaluations, with 3 surgeons using a 24-point scale.

Several authors used multiple aspects of evaluation: From and colleagues[81] rated airway management skills from 4 (excellent) to 1 (inadequate) for overall mask airway skill, overall intubation skill, tooth pressure, initial tube placement and efficient ventilation after placement. Evaluators classified patients from 1 (easy to intubate) to 4 (difficult to intubate) and from Mallampati class 1 (soft palate, fauces, uvula, pillars visible) to class IV (soft palate not visible). Carr and colleagues[78] assessed knowledge of the technique of anterior nasal packing with a 17-item short answer test and performance on a model assessed by a 16-item checklist; Matsumoto and colleagues[79] measured skills in removal of a midureteral stone with a semirigid ureteroscope and a basket by a global rating scale, a checklist, a pass rating and the time needed to complete the task. Summers and colleagues[86] rated each suturing technique on instrument handling, body position, accuracy, tightness, alignment and time to perform. Students also took a 50-item multiple choice test.

Only 2 authors used evaluative terms to describe the students' per-

Table 3

Assessment of the methodological quality of the included RCTs

	Selection bias		Performance bias			Attrition bias	Detection bias			
			Blinding							
RCT	Randomized	Randomization concealed	Participants	Instructors	Process analysis	Attrition analysis	Assessment blinded	Intention to treat	Power computation	Appropriate analysis
Abe et al[76]	Y	N	N	N	Y	All completed	N	All completed	N	Y
Ali et al[77]	Y	N	N	N	N	All completed	N	All completed	N	Y
Carr et al[78]	Y	N	N	N	Y	All completed	N	All completed	Y†	N‡
Engum et al[87]	N	N	N	N	Y	All completed	N	All completed	N	Y
From et al[81]	N	Y	N	Y	Y	All completed	Y	All completed	N	N‡
Gilbart et al[83]	N	N	N	N	Y	All completed	N	All completed	N	N‡
Hong et al[82]	N	N	N	N	Y	All completed	Y	All completed	N	Y
Jun et al[89]	N	N	N	N	Y	All completed	N	All completed	N	Y
Mann et al[91]	N	N	N	N	Y	All completed	N	All completed	N	Y
Matsumoto et al[79]	Y	N	N	Y	Y	All completed	Y	All completed	Y	Y
Petroianu et al[90]	N	N	N	N	Y	All completed	N	All completed	N	Y
Rogers et al[95]	N	N	N	Y	Y	All completed	N	All completed	N	Y
Rogers et al[93]	N	N	N	N	Y	108 at baseline; 105 completed	N	108 at baseline; 105 completed	N	Y
Rogers et al[84]	N	N	N	N	Y	All completed	N	All completed	N	N‡
Rogers et al[85]	N	N	N	N	Y	91 at baseline, 82 completed	N	91 at baseline, 82 completed	N	N‡
Rogers et al[94]	N	N	N	N	Y	All completed	N	27/40 videotapes useable	N	Y
Summers et al[86]	N	N	N	N	Y	All completed	N	All completed	N	N‡
Talan et al[88]	N	N	N	N	Y	All completed	N	All completed	N	Y
Todd et al[80]	Y	Y	N	Y	Y	89/91 completed	Y	89/91 completed	N	N‡

RCT = randomized controlled trial; Y = yes; N = no.
Although no study planned a process analysis, they were conducted de facto in most studies because the surgeons either instructed or observed and then rated the students.
†Post hoc.
‡Students were allocated to small groups where interaction was possible, and the effects of clustering were not allowed for in the statistical analysis.

formance: Todd and colleagues[80] described intubations as not competent to outstanding, and From and colleagues[81] described airway management skills as inadequate to excellent. It was not the stated purpose of any author to set standards for undergraduate achievement but, rather, to find more efficient ways to teach undergraduates. The setting of ideal and minimal safe scores is a topic for further research.

Overall risk of bias

In assessing the overall risk of bias in this group of 19 RCTs, we estimated that those studies that obtained significant results (thus not subject to type II error) are at low risk if the study randomized participants to individual tasks rather than to clusters and at moderate risk if clustering in the sample could have contributed to the outcomes and the effects of clustering were not adjusted for.

There were 2 RCTs of IV access. Engum and colleagues[87] found no differences between groups who learned on a computer simulator and by self-study, and Talan and colleagues[88] found no differences in success at IV access by cutdown at the cephalic or saphenous vein.

There were 2 RCTs of interosseous access. Abe and colleagues[76] found IV access in neonates was faster with the intraosseous route than the umbilical vein. Jun and colleagues[89] found that, after training, intraosseous access was more successful with the Cook Sur-Fast screw-tipped needle than with a standard bone marrow needle.

Because most of the outcomes were performance times, the absence of concealment and blinding might not have biased these results. None of the RCTs performed a power computation. The absence of significant results and of a power computation in Engum and colleagues[87] and Talan and colleagues[88] place these 2 RCTs at moderate risk of bias from type II error.

There were 2 RCTs of airway management. From and others[81] found no differences in success at intubating patients undergoing general anesthesia between the groups who took the American Heart Association self-study course and those who received a 1-hour lecture by anesthesiologists. Petroianu and others[90] found no differences with the Trachlight between the nasal or oral route in time to intubation or success by the tenth attempt to intubate. The absence of significant results and a power computation in From and colleagues[81] and Petroianu and colleagues,[90] and failure to allow for a clustered design in From and colleagues[81] place both at moderate risk of bias.

There was 1 RCT of CPR. Todd and others[80] found that those who took a 34-minute self-instruction CPR video were more competent when tested 100 days later than were those who took the American Heart Association Citizen Heartsaver CPR course A. The absence of statistical allowance for the clustered design places Todd and colleagues[80] at moderate risk of bias.

There are 2 RCTs of trauma assessment. Ali and colleagues[77] found that students who took the TEAM trauma evaluation and management program had a statistically significant improvement in scores, compared with the control groups ($p < 0.0001$) on the day of instruction, but no longer-term follow-up was undertaken. Gilbart and others[83] found no differences on trauma or nontrauma OSCE scores between groups who received computer and seminar instruction. Because of the absence of significant results, the absence of a power computation and the failure to statistically adjust for the cluster design, the study by Gilbart and others[83] is at moderate risk of bias.

There were 10 RCTs of surgical technique. Carr and colleagues[78] found that the groups that received instruction on anterior nasal packing for epistaxis by self-instruction on a computer or by face-to-face seminars had higher scores, compared with their pretests ($p < 0.05$). Hong and others[82] found that students who viewed interactive anatomy software about 2 operations they were to witness had higher test scores. Mann and others[91] found both groups of students who were taught about the Whipple procedure, either by watching a video or by cutting out paper shapes, improved in their understanding of anatomic relations, but there were no differences between the groups. The validity of the questionnaire used in determining test scores was not tested, but the authors had piloted the approach in a previous study of inguinal hernia repair.[92]

Matsumoto and colleagues[79] found that groups that practised ureteric stone extraction with either an inexpensive or an expensive (high fidelity) model supervised by a urologist had significantly higher scores than a group that received didactic instruction. Rogers and colleagues[94] found that students who studied clinical scenarios before their surgical intensive care elective improved their average test scores ($p < 0.0001$) on 3 of the OSCE stations (intubation, ventilator, hypotension) but not on the stations for airway, breathing and circulation or pulmonary artery data.

Rogers and others[95] tested several methods of teaching students to tie knots and found that a) a group that was shown the correct method and errors in tying knots improved their scores ($p < 0.01$), but groups shown no errors and only correct methods and those shown only errors did not improve[95]; b) students who took a computer-assisted teaching session with individualized feedback from surgical faculty had greater improvement than did those who only received computer instruction ($p < 0.001$)[93]; c) students who practised in groups of 6 to 8 students, with each using an individual computer, and those in pairs using 1 computer and giving each other feedback both improved in the proportion of correctly tied knots ($p < 0.001$). The authors concluded that peers could not substitute for expert faculty in this exercise.[84] A computer and a lecture group had no significant differences in the proportions of correctly tied square knots or in the average time per knot.[85] Summers and colleagues[86] compared computer, videotape and didactic groups and found that the computer group had more complete knots at both immediate ($p < 0.01$) and 1-month follow up ($p < 0.01$) than did the didactic or videotape groups, and the didactic group scored higher on MCQs both times.

Because of the absence of significant results and failure to conduct a power computation, the study by Mann and colleagues[91] is at moderate risk of bias. Because of the failure to correct statistically for the clustered design, 5 others (Carr et al,[78] Rogers et al[84,85] Summers et al,[86] Todd et al[80]) are at moderate risk of bias.

Discussion

The interventions in this review were developed by practising academic surgeons to enhance teaching and learning in existing rotations. The interventions were thoughtfully designed, but the execution of some of the trials exposed them to the risk of bias. It was not the purpose of the researchers to establish ideal or safe minimum scores. It remains for further research to establish whether more instruction, more practice or both would enable students to achieve higher average scores and, presumably, greater procedural proficiency.

Norman[96] notes that controlled trials in medical education often have a large unexplained variance in their results. He advocates the model of psychological research whereby the circumstances of the experiment are tightly controlled, and factors that might contribute to the results are systematically varied over a series of experiments based on a theory of causation, so that the effective causes are understood. This model could be applied to this field of study.

Educators and researchers could correct the design and execution problems noted in these RCTs. Improvements could include making a power computation; concealing randomization by blinding students, instructors and assessors; evaluating the different components of the interventions to assess which need strengthening; psychometric analysis of outcome measures to optimize their reliability and validity; assessing which aspects of instruction improve outcomes; describing and correcting problems in the learning process; identifying which students have difficulty with which aspects of the learning process; and correcting for the effects of clustering in the statistical analysis.

This review has described approaches to managaeble research questions (i.e., how to improve and measure the ability of students to tie square knots), but there are no reported RCTs on many common surgical procedures, such as foreign body extraction, wound assessment, common fractures or burns. Communication between surgical and primary care program directors could begin a planning process to conceptualize how to improve the learning of surgical and emergency skills and to identify the high-priority procedural skills requiring educational research and a specific curriculum. Consortia of programs or medical schools could cooperate to achieve the larger sample sizes.

Conclusion

RCTs of teaching undergraduates surgical and emergency skills have many positive study outcomes, but there are significant methodological weaknesses in study design. This systematic review provides a baseline for further work important to both medical education and clinical practice.

Competing interests: None declared.

References

1. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; 354:1896-900.
2. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6.* In: The Cochrane Library, Chichester (UK): John Wiley and Sons; 2006.
3. Cázares A, Beatty LA. Scientific methods for prevention intervention research. NIDA research monograph 139. Rockville (MD): National Institute on Drug Abuse; 1994. p. 115-26.
4. Murray DM, Hannan PJ. Planning for the appropriate analysis in school-based drug-use prevention studies. *J Consult Clin Psychol* 1990;58:458-68.
5. Murray DM, Short BJ. Intraclass correlation among measures related to tobacco use by adolescents: estimates, correlates, and applications in intervention studies. *Addict Behav* 1997;22:1-12.
6. Palmer RF, Graham JW, White EL, et al. Applying multilevel analytic strategies in adolescent substance use prevention research. *Prev Med* 1998;27:328-36.
7. Ali J, Adam R, Williams JI, et al. Teaching effectiveness of the trauma evaluation and management module for senior medical students. *J Trauma* 2002;52:847-51.
8. Ali J, Cohen RJ, Gana TJ, et al. Effect of the Advanced Trauma Life Support program on medical students' performance in simulated trauma patient management. *J Trauma* 1998;44:588-91.
9. Ali I, Cohen R, Reznick R. Demonstration of acquisition of trauma management skills by senior medical students completing the ATLS program. *J Trauma* 1995; 38:687-91.
10. Guarino JR, Guarino JC. Auscultatory percussion: A simple method to detect pleural effusion. *J Gen Intern Med* 1994;9:71-4.
11. Gunther SB, Soto GE, Colman WW. Interactive computer simulations of knee-replacement surgery. *Acad Med* 2002;77:753-4.
12. Hanna GB, Creesswell AB, Cuschieri A. Shadow depth cues and endoscopic task performance. *Arch Surg* 2002;137:1166-9.
13. Kline P, Shesser R, Smith M, et al. Comparison of a videotape instructional program with a traditional lecture series for medical student emergency medicine teaching. *Ann Emerg Med* 1986;15:16-8.
14. Liddell MJ, Davidson S K, Taub H, et al. Evaluation of procedural skills training in an undergraduate curriculum. *Med Educ*

2002;36:1035-41.
15. Lossing A, Groetzsch G. A prospective controlled trial of teaching basic surgical skills with 4th year medical students. *Med Teach* 1992;14:49-52.
16. Lucas J, McKay S, Baxley E. EKG arrhythmia recognition: a third-year clerkship teaching experience. *Fam Med* 2003;35:163-4.
17. Lunenfeld E, Weinreb B, Lavi Y, et al. Assessment of emergency medicine: a comparison of an experimental objective structured clinical examination with a practical examination. *Med Educ* 1991;25:38-44.
18. Macmillan CSA, Crosby JR, Wildsmith JAW. Skilled task teaching and assessment. *Med Teach* 2001;23:591-4.
19. Mehrabi A, Glückstein CH, Benner A, et al. A new way for surgical education–development and evaluation of a computer-based training module. *Comput Biol Med* 2000;30:97-109.
20. Owen SG, Hall R, Anderson J, et al. Programmed learning in medical education. *Postgrad Med J* 1965;41:201-5.
21. Reid JDS, Vestrup JA. Use of a simulation to teach central venous access. *J Med Educ* 1988;63:196-7.
22. Rogers DA, Elstein AS, Bordage G. Improving continuing medical education for surgical techniques: applying the lessons learned in the first decade of minimal access surgery. *Ann Surg* 2001;233:159-66.
23. Rogers PL, Jacob H, Rashwan AS, et al. Quantifying learning in medical students during a critical care medicine elective: a comparison of three evaluation instruments. *Crit Care Med* 2001;29:1268-73.
24. Seraj MA, Naguib M. Cardiopulmonary resuscitation skills of medical professionals. *Resuscitation* 1990;20:31-9.
25. Spielman FJ, Murphy CA Jr, Levin KJ. Medical student education in life-support skills. *J Med Educ* 1983;58:637-40.
26. Tandberg D, Kastendieck KD, Meskin S. Observer variation in measured ST-segment elevation. *Ann Emerg Med* 1999;34:448-52.
27. Vogelgesang SA, Karplus TM, Kreiter CD. An instructional program to facilitate teaching joint/soft-tissue injection and aspiration. *J Gen Intern Med* 2002;17:441-5.
28. Ahlberg G, Heikkinen T, Iselius L, et al. Does training in a virtual reality simulator improve surgical performance? *Surg Endosc* 2002;16:126-9.
29. Aliabadi-Wahle S, Ebersole M, Choe EU, et al. Training in clinical breast examination as part of a general surgery core curriculum. *J Cancer Educ* 2000;15:10-3.
30. Ault GT, Sullivan M, Chalabian J, et al. A focused breast skills workshop improves the clinical skills of medical students. *J Surg Res* 2002;106:303-7.
31. Carpenter LG, Piermattei DL, Salman MD, et al. A comparison of surgical training with live anesthetized dogs and cadavers. *Vet Surg* 1991;20:373-8.
32. Dorafshar AH, O'Boyle DJ, McCloy RF. Effects of a moderate does of alcohol on simulated laparoscopic surgical performance. *Surg Endosc* 2002;16:1753-8.
33. Dunnington G, Witzke D, Rubeck R, et al. A comparison of the teaching effectiveness of the didactic lecture and the problem-oriented small group session: a prospective study. *Surgery* 1987;102:291-6.
34. Erkonen WE, Krachmer M, Cassell MD, et al. Cardiac anatomy instruction by ultrafast computed tomography versus cadaver dissection. *Invest Radiol* 1992;27:744-7.
35. Fan KH, Hung OR, Agro F. A comparative study of tracheal intubation using an intubating laryngeal mask (Fastrach) alone or together with a lightwand (Trachlight). *J Clin Anesth* 2000;12:581-5.
36. Fincher RE, Abdulla A M, Sridharan MR, et al. Computer-assisted learning compared with weekly seminars for teaching fundamental electrocardiography to junior medical students. *South Med J* 1988;81:1291-4.
37. Fincher RM, Abdulla A M, Sridharan MR, et al. Comparison of computer-assisted and seminar learning of electrocardiogram interpretation by third-year students. *J Med Educ* 1987;62:693-5.
38. Fincher RM, Abdulla A M, Sridharan MR, et al. Teaching fundamental electrocardiography to medical students: computer-assisted learning compared with weekly seminars. *Res Med Educ* 1987;26:197-202.
39. Fincher RM, Abdulla AM, Sridharan MR, et al. A prospective educational trial comparing efficacy of computer-assisted learning and weekly seminars in teaching EKG interpretation. *Res Med Educ* 1986;25:3-7.
40. Grum CM, Gruppen LD, Woolliscroft JO. The influence of vignettes on EKG interpretation by third-year students. *Acad Med* 1993;68(Suppl 10):S621-S623
41. Halm B, Yamamoto LG. Comparing ease of intraosseous needle placement: Jamshidi versus Cook. *Am J Emerg Med* 1998;16:420-2.
42. Hassett J, Luchette F, Doerr R, et al. Utility of an oral examination in a surgical clerkship. *Am J Surg* 1992;164:372-6.
43. Hatala RM, Brooks LR, Norman GR. Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation. *Adv Health Sci Educ Theory Pract* 2003;8:17-26.
44. Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med* 1999;14:126-9.
45. Hill DA, Lord DSA. Complementary value of traditional bedside teaching and structured clinical teaching in introductory surgical studies. *Med Educ* 1991;25:471-4.
46. Holmes JF, Panacek EA, Sakles JC, et al. Comparison of 2 cricothyrotomy techniques: Standard method versus rapid 4-step technique. *Ann Emerg Med* 1998;32:442-6.
47. Holtedahl KA, Bø B, Hansen AH, et al. Studentundervisning i allmennmedisinsk ferdighetslaboratorium. *Tiddskr Nor Lægeforen* 1999;19:2854-7.
48. Hung OR, Pytka S, Morris I, et al. Clinical trial of a new lightwand device (Trachlight) to intubate the trachea. *Anesthesiology* 1995;83:509-14.
49. Hyltander A, Liljegren E, Rhodin PH, et al. The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surg Endosc* 2002;16:1324-8.
50. Jonas JB, Rabethge S, Bender HJ. Computer-assisted training system for pars plana vitrectomy. *Acta Ophthalmol Scand* 2003;81:600-4.
51. Jones DB, Brewer JD, Soper NJ. The influence of three-dimensional video systems on laparoscopic task performance. *Surg Laparosc Endosc* 1996;6:191-7.
52. Kim JH, Kim WO, Min KT, et al. Learning by computer simulation does not lead to better test performance than textbook study in the diagnosis and treatment of dysrhythmias. *J Clin Anesth* 2002;14:395-400.
53. Kothari SN, Kaplan BJ, DeMaria EJ, et al. Training in laparoscopic suturing sklills using a new computer-based virtual reality simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. *J Laparoendosc Adv Surg Tech A* 2002;12:167-73.
54. Kovacs G, Bullock G, Ackroyd-Stolarz S, et al. A randomized controlled trial on the effect of educational interventions in promoting airway management skill maintenance. *Ann Emerg Med* 2000;36:301-9.
55. Kurola J, Harve H, Kettunen T, et al. Airway management in cardiac arrest–comparison of laryngeal tube, tracheal intubation and bag-valve mask ventilation in emergency medical training. *Resuscitation* 2004;61:149-53.
56. Lee SK, Pardo M, Gaba D, et al. Trauma assessment training with a patient simula-

tor: a prospective, randomized study. *J Trauma* 2003;55:651-7.

57. Livingstone RA, Ostrow DN. Professional patient-instructors in the teaching of the pelvic examination. *Am J Obstet Gynecol* 1978;132:64-7.
58. Madan AK, Alibadi-Wahle S, Babbo AM, et al. Education of medical stuents in clinical breast examination during surgical clerkship. *Am J Surg* 2002;184:637-41.
59. Mangione S, Nieman LZ, Greenspon LW, et al. A comparison of computer-assisted instruction and small group teaching of cardiac auscultation to medical students. *Med Educ* 1991;25:389-95.
60. Mangione S, Nieman LZ, Gracely EJ. Comparison of computer-based learning and seminar teaching of pulmonary auscultation to first-year medical students. *Acad Med* 1992;67(Suppl 10):S63-S65.
61. Nio D, Bemelman WA, den Boer KT, et al. Efficiency of manual vs robotical (Zeus) assisted laparoscopic surgery in the performance of standardized tasks. *Surg Endosc* 2002;16:412-5.
62. Ostrow DN, Craven N, Cherniack RM. Learning pulmonary function interpretation: Deductive versus inductive methods. *Am Rev Respir Dis* 1975;112:89-92.
63. Pearson AM, Gallagher AG, Rosser JC, et al. Evaluation of structured and quantitative methods for teaching intracorporeal knot tying. *Surg Endosc* 2002;16:130-7.
64. Perkins GD, Hulme J, Bion JF. Peer-led resuscitation training for healthcare students: a randomised controlled study. *Intensive Care Med* 2002;28:698-700.
65. Rogers PL, Grenvik A, Willenkin RL. Teaching medical students complex cognitive skills in the intensive care unit. *Crit Care Med* 1995;23:575-81.
66. Schwind CJ, Boehler ML, Folse R, et al. Development of physical examination skills in a third-year surgical clerkship. *Am J Surg* 2001;181:338-40.
67. Shatzer JH, Darosa D, Colliver JA, et al. Station length requirements for reliable performance-based examination scores. *Acad Med* 1993;68:224-9.
68. Shockley LW, Butzier DJ. A modified wire-guided technique fore venous cutdown access. *Ann Emerg Med* 1990;19:393-5.
69. Ström P, Kjellin A, Hedman L, et al. Training in tasks with different visual-spatial components does not improve virtual arthroscopy performance. *Surg Endosc* 2004;18:115-20.
70. Sullivan ME, Ault GT, Hood DB, et al. The standardized vascular clinic: an alternative to the traditional ambulatory setting. *Am J Surg* 2000;179:243-6.
71. Sutyak JP, Lebeau RB, O'Donnell AM. Unstructured cases in case-based learning benefit students with primary care career preferences. *Am J Surg* 1998;175:503-7.
72. Sutyak JP, Lebeau RB, Spotnitz AJ, et al. Role of case structure and prior experience in a case-based surgical clerkship. *Am J Surg* 1996;172:286-90.
73. Torkington J, Smith SGT, Rees BI, et al. Skill transfer from virtual reality to a real laparoscopic task. *Surg Endosc* 2001;15:1076-9.
74. Vontver L, Irby D, Rakestraw P, et al. The effects of two methods of pelvic examination instruction on student performance and anxiety. *J Med Educ* 1980;55:778-85.
75. Wilhelm DM, Ogan K, Roehrborn CG, et al. Assessment of basic endoscopic performance using a virtual reality simulator. *J Am Coll Surg* 2002;195:675-81.
76. Abe KK, Blum GT, Yamamoto LG. Intraosseous is faster and easier than umbilical venous catheterization in newborn emergency vascular access models. *Am J Emerg Med* 2000;18:126-9.
77. Ali J, Danne P, McColl G. Assessment of the trauma in evaluation management (TEAM) module in Australia. *Injury* 2004;35:753-8.
78. Carr MM, Reznick RK, Brown DH. Comparison of computer-assisted instruction and seminar instruction to acquire psychomotor and cognitive knowledge of epistaxis management. *Otolaryngol Head Neck Surg* 1999;121:430-4.
79. Matsumoto ED, Hamstra SJ, Radomski SB, et al. The effect of bench model fidelity on endourological skills: A randomized controlled study. *J Urol* 2002;167:1243-7.
80. Todd KH, Braslow A, Brennan RT, et al. Randomized controlled trial of video self-instruction versus traditional CPR training. *Ann Emerg Med* 1998;31:364-9.
81. From RP, Pearson KS, Albanese MA, et al. Assessment of an interactive learning system with “sensorized” mannequin for airway management instruction. *Anesth Analg* 1994;79:136-42.
82. Hong D, Regehr G, Reznick RK. The efficacy of a computer-assisted preoperative tutorial for clinical clerks. *Can J Surg* 1996;39:221-4.
83. Gilbart MK, Hutchinson CR, Cusimano MD, et al. A computer-based trauma simulator for teaching trauma management skills. *Am J Surg* 2000;179:223-8.
84. Rogers DA, Regehr G, Gelula M, et al. Peer teaching and computer-assisted learning: An effective combination for surgical skill training? *J Surg Res* 2000;92:53-5.
85. Rogers DA, Regehr G, Yeh KA, et al. Computer-assisted learning versus a lecture and feedback seminar for teaching a basic surgical technical skill. *Am J Surg* 1998;175:508-10.
86. Summers AN, Rinehart GC, Simpson D, et al. Acquisition of surgical skills: a randomized trial of didactic, videotape, and computer-based training. *Surgery* 1999;126:330-6.
87. Engum SA, Jeffries P, Fisher L. Intravenous catheter training system: Computer-based education versus traditional learning methods. *Am J Surg* 2003;186:67-74.
88. Talan DA, Simon RR, Hoffman JR. Cephalic vein cutdown at the wrist: comparison to the standard saphenous vein ankle cutdown. *Ann Emerg Med* 1988;17:38-42.
89. Jun H, Haruyama AZ, Chang KSG, et al. Comparison of a new screw-tipped intraosseous needle versus a standard bone marrow aspiration needle for infusion. *Am J Emerg Med* 2000;18:135-9.
90. Petroianu GA, Subotic S, Heil P, et al. Intubation with transillumination: nasal or oral? *Prehospital Disaster Med* 1999;14:104-6.
91. Mann BD, Heath CM, Gracely E, et al. Use of a paper-cut as an adjunct to teaching the Whipple procedure by video. *Am J Surg* 1998;176:379-83.
92. Mann BD, Seidman A, Haley T, et al. Teaching three-dimensional surgical concepts of inguinal hernia using a time-effective manner using a two-dimensional paper-cut. *Am J Surg* 1997;173:542-5.
93. Rogers DA, Regehr G, Howdieshell TR, et al. The impact of external feedback on computer-assisted learning for surgical technical skill training. *Am J Surg* 2000;179:341-3.
94. Rogers PL, Jacob H, Thomas EA, et al. Medical students can learn the basic application, analytic, evaluative and psychomotor skills of critical care medicine. *Crit Care Med* 2000;28:550-4.
95. Rogers DA, Regehr G, MacDonald J. A role for error training in surgical technical skill instruction and evaluation. *Am J Surg* 2002;183:242-5.
96. Norman G. RCT = results confounded and trivial: the perils of grand educational experiments. *Med Educ* 2003;37:582-4.