REVIEWS

Patient Outcomes in Simulation-Based Medical Education: A Systematic Review

Benjamin Zendejas, MD, MSc[1], Ryan Brydges, PhD[2], Amy T. Wang, MD[3], and David A. Cook, MD, MHPE[3,4]

[1]Department of Surgery, Mayo Clinic College of Medicine, Rochester, MN, USA; [2]Department of Medicine, University of Toronto, Toronto, ON, Canada; [3]Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; [4]Office of Education Research, Mayo Medical School, Rochester, MN, USA.

OBJECTIVES: Evaluating the patient impact of health professions education is a societal priority with many challenges. Researchers would benefit from a summary of topics studied and potential methodological problems. We sought to summarize key information on patient outcomes identified in a comprehensive systematic review of simulation-based instruction.

DATA SOURCES: Systematic search of MEDLINE, EMBASE, CINAHL, PsychINFO, Scopus, key journals, and bibliographies of previous reviews through May 2011.

STUDY ELIGIBILITY: Original research in any language measuring the direct effects on patients of simulation-based instruction for health professionals, in comparison with no intervention or other instruction.

APPRAISAL AND SYNTHESIS: Two reviewers independently abstracted information on learners, topics, study quality including unit of analysis, and validity evidence. We pooled outcomes using random effects.

RESULTS: From 10,903 articles screened, we identified 50 studies reporting patient outcomes for at least 3,221 trainees and 16,742 patients. Clinical topics included airway management (14 studies), gastrointestinal endoscopy (12), and central venous catheter insertion (8). There were 31 studies involving postgraduate physicians and seven studies each involving practicing physicians, nurses, and emergency medicine technicians. Fourteen studies (28 %) used an appropriate unit of analysis. Measurement validity was supported in seven studies reporting content evidence, three reporting internal structure, and three reporting relations with other variables. The pooled Hedges' g effect size for 33 comparisons with no intervention was 0.47 (95 % confidence interval [CI], 0.31–0.63); and for nine comparisons with non-simulation instruction, it was 0.36 (95 % CI, −0.06 to 0.78).

LIMITATIONS: Focused field in education; high inconsistency (I^2>50 % in most analyses).

CONCLUSIONS: Simulation-based education was associated with small-moderate patient benefits in comparison with no intervention and non-simulation instruction, although the latter did not reach statistical significance. Unit of analysis errors were common, and validity evidence was infrequently reported.

KEY WORDS: medical education; outcomes research; simulation; educational technology; program evaluation; quantitative research methods.

J Gen Intern Med 28(8):1078–89

DOI: 10.1007/s11606-012-2264-5

INTRODUCTION

Evaluating the patient impact of health professions education has become a societal priority in the past decade,[1,2] concurrent with increased emphasis on evidence-based medicine, patient safety, and practice efficiencies. Although there are legitimate concerns about the use of patient outcomes,[3] few would argue against the appeal of measuring patient effects as "translational" outcomes of medical education.[4] Systematic reviews indicate that patient outcomes are reported in 0–5 % of medical education studies.[5–9] However, such studies are at risk of being over-interpreted. The validity of the measurements and the integrity of the statistical analyses are particularly important, for if measurements are invalid or statistical analyses are flawed, the results cannot be trusted.

Researchers aspiring to measure patient outcomes would benefit from knowing the clinical topics that have and have not been studied, the methodological problems that should be avoided, and the magnitude of expected effects. The purpose of our study was to fill these gaps, using data from a comprehensive systematic review of simulation-based instruction.

Technology-enhanced simulation has emerged as a powerful tool in health professions education.[10–12] Previous reports from our review have demonstrated that simulation is superior to no intervention (609 studies)[13] and to non-simulation instruction (92 studies),[14] and have used comparisons of different simulation interventions (289 studies) to identify evidence-based best practices.[15] The present study is a planned sub-analysis of these data. We aimed to

Electronic supplementary material *The online version of this article (doi:10.1007/s11606-012-2264-5) contains supplementary material, which is available to authorized users.*

Published online April 18, 2013

summarize studies evaluating patient outcomes, identifying the clinical areas addressed, the methodological strengths and flaws common to such research, and the magnitude of effect that might be expected in such studies. We are not aware of such a review published in any field of medical education.

Methodological Issues in Patient Outcomes Studies

In this review, we focused on two methodological issues: measurement validity and statistical integrity. Measurement validity can be evaluated by accruing evidence from five sources: 1) content (steps to ensure that the instrument reflects what it is intended to measure); 2) internal structure (reproducibility or factor structure); 3) relations with other variables (associations with another measure or with training level); 4) response process (analysis of rater response or test security); and 5) consequences (the downstream impact of the assessment itself).[16,17] Systematic reviews have documented that validity evidence is infrequently reported. Estimates vary widely depending on the sample, but content, internal structure, and relations with other variables evidence are typically reported in < 40 % of studies,[6,9,13,18–21] and response process and consequences are reported in < 10 %.[18] However, the reporting of validity evidence for patient outcomes is unknown.

In an education intervention study, the unit of intervention is the trainee, and thus the unit of statistical analysis should also be the trainee, not the patient.[22] Studies in which multiple patients contribute data for each trainee (patients *clustered* in trainees) require statistical techniques that account for such clustering.[23] Clinical research suggests that such *unit of analysis errors* are relatively common, ranging from 22 to 71 %,[24–29] and generally inflate the power of the analysis.[26,30] This may lead to conclusions of statistical significance when none are warranted. We are not aware of studies evaluating the prevalence of unit of analysis error in health professions education. The CONSORT extension for cluster randomized trials[31] also requests information on how clustering was incorporated into sample size calculations, and how clusters and individuals (i.e., trainees and patients) progress through the trial. During our review, we further noted frequent independent analysis of multiple similar outcomes (multiple independent hypothesis testing). We sought to determine the prevalence of these methodological deficiencies.

METHODS

This review is a planned sub-analysis of a comprehensive systematic review of technology-enhanced simulation; more detailed methods have been published previously.[13] It was planned, conducted, and reported in adherence to current standards of quality for reporting systematic reviews.[32]

Questions

We sought to evaluate the type, clinical task, and average effect size of patient outcomes in simulation-based education. For each outcome, we also sought to evaluate the frequency of validity evidence reporting and quality of statistical analysis.

Study Eligibility and Definitions

We included all comparative studies that used patient outcomes to evaluate technology-enhanced simulation for training health professionals in any field or stage of training. Health professionals included student and practicing physicians, nurses, emergency medicine technicians, and other allied health providers. We included studies making comparison with no intervention (i.e., a control arm or pre-intervention assessment) or alternate instruction. We made no exclusions based on language or year of publication.

We defined technology-enhanced simulation as an educational tool or device with which the learner physically interacts to mimic an aspect of clinical care.[13] Computer-based virtual patients and human patient actors (standardized patients) did not qualify as technology-enhanced simulation, but did count as comparison interventions.

We defined patient outcome as a direct effect on a patient, such as complication or procedural success, in contrast with trainee behaviors such as proficiency or efficiency, which may or may not have the desired effect (e.g., a technically poor performance may still have a good outcome, and conversely a complication may follow a technically correct performance). We further classified patient effects as those that happen *to* the patient but may not affect morbidity or mortality (e.g., procedural success or delay in diagnosis), and those that arise *within* the patient (e.g., survival or complications; see Table 1).

Study Identification

With the assistance of a research librarian, we searched multiple databases including MEDLINE, EMBASE, CINAHL, PsychINFO, ERIC, Web of Science, and Scopus. We also examined the reference lists of key review articles and 190 included articles, and the full table of contents of two journals devoted to health professions simulation. The last search date was May 11, 2011. This search strategy has been published in full.[13]

Study Selection

To identify studies for inclusion, two reviewers independently screened the titles and abstracts of all potentially

Table 1. Patient Outcomes Reported in Simulation-Based Education Research

Outcome (n*)	Examples (antecedent event)
Within-patient outcomes: conditions or events that arise from or within the patient	
Complications (24)	Bloodstream infection (central line placement)[50]
	Pneumothorax (thoracentesis)[54]
	Perforation (colonoscopy)[46]
Patient discomfort during event (7)	Patient discomfort (colonoscopy)[39]
Survival (6)	Survival to discharge (cardiac resuscitation)[48]
	Stillbirth (obstetric delivery with umbilical cord prolapse)[57]
Duration of stay (2)	Duration of hospitalization (cardiac resuscitation)[48]
Patient satisfaction (2)	Patient satisfaction (intrauterine device insertion)[73]
Patient symptoms / quality of life (0)†	(none found in this sample)
Laboratory test results (0)†	(none found in this sample)
Patient compliance (0)†	(none found in this sample)
Patient motivation (0)†	(none found in this sample)
To-patient outcomes: conditions or events that happen to the patient	
Procedural success (31)	Successful endotracheal intubation[72]
	Reach cecum (colonoscopy)[44]
	Successful venous cannulation[83]
Evaluation of final product (2)	Tissue removed during transurethral resection of prostate[67]
Accuracy of diagnosis (1)	Major pathology identified (upper gastrointestinal endoscopy)[66]
Delay in diagnosis (1)	Time to computed tomography (CT) scan (major trauma)[65]
Delay in critical action (1)	Time to operating room (major trauma)[65]

**Number of studies reporting one or more outcomes of this type (many studies reported > 1 outcome)*
†Identified in advance as potential outcomes; none identified in this sample, but included here for completeness of the model

eligible articles. For articles that could not be excluded based on title/abstract, we obtained and reviewed the full text, again independently and in duplicate. We resolved all disagreements by consensus. Chance-adjusted interrater agreement, determined using the intraclass correlation coefficient (ICC), was 0.69.

Data Extraction

We abstracted information from each study using a standardized abstraction form. Two independent reviewers abstracted all information for which reviewer judgment was required, with disagreements resolved by consensus. ICC for identification of patient outcomes (vs. other study outcomes) was 0.74. We coded the number and type of patient outcomes (ICC 0.84), and further classified these as within-patient or to-patient events (ICC 1.0).

We abstracted information on study methods, including outcome validity evidence, study design, method of group assignment, and blinding of assessments, using the Medical Education Research Study Quality Instrument[6] (MERSQI) and an adaptation of the Newcastle-Ottawa Scale (NOS) for cohort studies.[33] We coded additional methodological issues specific to our research questions, including:

- the unit of analysis (patient or trainee; ICC 0.61),
- whether patient outcomes were linked to the trainee or reported in aggregate (ICC 0.83),
- the data source (trainee, patient, investigator, or patient record; ICC 0.99),
- a priori power calculation reported (ICC 0.84),
- patient demographic information reported (ICC .89),
- patient outcome identified as the primary (vs. secondary) outcome (ICC 0.70), and
- patient outcome prespecified (i.e., listed as a planned outcome in the study objective or methods; ICC 0.46 with raw agreement 92 %).

Data Synthesis

We synthesized outcomes quantitatively using random-effects meta-analysis. We first calculated a standardized mean difference (Hedges' g effect size) using methods described previously.[13] For articles reporting insufficient information to calculate an effect size, we requested additional information from authors. We conducted separate meta-analyses for studies making comparison with a) no intervention, b) non-simulation instruction, and c) another simulation-based instructional intervention. For all analyses, we planned sensitivity analyses excluding studies that used p-value upper limits or imputed standard deviations to estimate the effect size. We also planned subgroup analyses based on topic, study design (randomized versus nonrandomized and one-group pre-post vs. two-group), unit of analysis (trainee vs. patient), and blinding. We did not conduct subgroup analyses for the non-simulation and simulation–simulation comparisons, due to the paucity of studies.

For the simulation–simulation studies, we first coded each study arm for several key features of instructional design.[34] Then, for each feature we conducted a separate meta-analysis pooling the results of studies in which that feature varied between study arms (i.e., if a given feature were present equally in both arms, then that study would

not be included in the meta-analysis for that feature). This approach has been described in detail previously.[15]

The weighting for all meta-analyses was based on the number of trainees, not the number of patients. We quantified between-study inconsistency (heterogeneity) using the I^2 statistic,[35] which estimates the percentage of variability not due to chance. I^2 values > 50 % indicate large inconsistency. We used SAS 9.1 (SAS Institute, Cary, NC) for all analyses. Statistical significance was defined by a two-sided alpha of 0.05, and interpretations of clinical significance emphasized confidence intervals in relation to Cohen's effect size classifications (0.5–0.8=moderate, 0.2–0.5=small).[36]

RESULTS

Trial Flow

Of the 985 studies meeting initial inclusion criteria, 50 (5 %) reported one or more patient outcomes (see Appendix eFigure 1). These studies enrolled at least 3,221 trainees and reported data on over 16,742 patients. Of the 34 studies making comparison with no intervention,[37–70] 32 were included in our group's previous meta-analysis,[13] one was not included in that analysis due to missing information,[37] and for one,[62] we identified the patient outcome after publication of that review. In addition, we include in the present review nine studies making comparison with non-simulation instruction[71–79] and eight studies making comparison with alternate simulation.[71,80–86] These comparisons with active interventions were included in previous meta-analyses.[14,15] Three articles omitted the number of trainees; we contacted these authors and two provided needed information.

Study Features

Key information is summarized in Table 2. The first study was published in 1979, with only seven more studies published over the ensuing 22 years (see eFigure 2). By contrast, over half the studies were published in or after 2008. One study was published in French and one in Spanish. Thirty studies originated from the USA, ten from Europe, five from Asia, four from Canada, and one from Central America. The most common clinical topics were airway management (*N*=14), gastrointestinal endoscopy (*N*= 12), and central venous catheter insertion (*N*= 8). Over half the studies involved postgraduate physician trainees (i.e., residents; *N*=31), followed by practicing physicians, nurses, and emergency medicine technicians (seven studies each). Studies involving medical and nursing students were few (three each).

Most studies (31) reported one or more outcomes indicating procedural success (e.g., successfully reaching the cecum in a colonoscopy), while 24 reported complications (such as bloodstream infection or pneumothorax). Seven reported patient discomfort during a procedure, six reported survival, and two reported duration of hospitalization (see Table 1).

Study Quality: Participant Flow, Analysis Errors, Validity Evidence

Table 3 summarizes the methodological quality of included studies. Of the 50 studies, 47 articles reported the number of trainees enrolled and two authors supplied this information upon request. Among these 49 studies, the average enrollment was 65.7 trainees (median 34; range 5–300). Seven studies (of 50) did not report trainee follow-up (i.e., the number of trainees contributing to patient outcomes results). Forty studies reported the number of patients, with an average sample size of 419 (median 145; range 24–7,650). Among 36 studies providing information on both trainees and patients, the average number of patients per trainee ranged from 0.8 (i.e., the number of patients contributing information was fewer than the number of trainees) to 170, with a mean (median) 16.1 (3.8).

Three studies reported one patient observation per trainee. Among the remaining 47 studies, 29 used an inappropriate unit of analysis (i.e., failed to account for clustering), and an additional seven reported insufficient information to discern the unit of analysis. Patient data were linked with the care-providing trainee in 38 studies; in the remainder, aggregate patient data were analyzed.

Few studies reported validity evidence to support the interpretations of outcome measurements: seven provided content evidence, three provided internal structure evidence, and three reported relations with other variables. None reported response process or consequences evidence.

The patient outcome was identified as the primary outcome in 13 studies, and as a secondary outcome in five studies. The patient outcome was not mentioned in the objective or methods in three studies. A power statement for the patient outcome was present in seven studies. None of these power statements made mention of clustering or unit of analysis, although in one case there was a 1:1 relation between trainees and patients, so no adjustment was required. Four studies reported and adjusted analyses based on patient demographic data, 18 studies reported demographics without adjustment, and one study adjusted for demographics without reporting this information.

Quantitative Synthesis: Comparison with No Intervention

Thirty-four studies made comparison with no intervention (e.g., single-group pretest-posttest study, or comparison with a no-training arm), and 33 of these (with 1,694 trainees providing data) contained sufficient information to include in meta-analysis.[38–70] The pooled effect size (see Fig. 1) for these interventions was 0.47 (95 % confidence

Table 2. Studies Evaluating Patient Outcomes

Author (year)	Participants N; type	Study design	Comp.	Task	Outcome(s)	Unit of analysis
Lefcoe DL (1979)[71]	30; O	RCT	OE,Sim	Dental (tooth planing)	EVAL FINAL	Missing
Stewart RD (1984)[80]	122; EMT	Obs	Sim	Airway: intubation	SUCCESS COMPLICATIONS: trauma teeth, trauma lips, esophageal intubation, right mainstem intubation, prolonged trial, vomiting, other	Patient
Ovassapian A (1988)[72]	32; PG	RCT	OE	Airway: fiberoptic intubation	SUCCESS	Trainee
Stratton SJ (1991)[81]	125; EMT	RCT	Sim	Airway: intubation	SUCCESS COMPLICATIONS: esophageal intubation, dislodged tube, mainstem intubation, aspiration, oral trauma, dental trauma, leak	Trainee
Trooskin SZ (1992)[82]	26; EMT	RCT	Sim	Airway: intubation	SUCCESS	Trainee
Limpaphayom K (1997)[73]	300; O	Obs	OE	Intrauterine device insertion	SATISFACTION	Missing
Hosking EJ (1998)[60]	46; MS	Obs	NI	Airway: intubation	SUCCESS	Patient
Naik VN (2001)[77]	24; PG	RCT	OE	Airway: fiberoptic intubation	SUCCESS	Trainee
Chang KK (2002)[83]	28; NS,RN	RCT	Sim	Peripheral venous cannulation	SUCCESS	Trainee
Swanson ER (2002)[37]	n/a; RN,EMT	1PP	NI	Airway: intubation	SUCCESS	Patient
Gerson LB (2003)[74]	16; PG	Obs	OE	GI: sigmoidoscopy	SUCCESS: reach splenic flexure, retroflexion SATISFACTION: 3 satisfaction questions DISCOMFORT: 4 discomfort questions	Patient
Di Giulio E (2004)[38]	22; PG	RCT	NI	GI: upper gastrointestinal endoscopy	SUCCESS: completed [with, without assistance], esophageal intubation COMPLICATIONS: no complications	Patient
Rumball C (2004)[84]	81; EMT	Obs	Sim	Airway: intubation	SUCCESS COMPLICATIONS: inadequate ventilation, esophageal, hypopharyngeal, endobronchial intubation, trauma, incorrect port	Trainee
Sedlack RE (2004)[39]	8; PG	RCT	NI	GI: colonoscopy	DISCOMFORT	Patient
Sedlack RE (2004)[40]	38; PG,MD	RCT	NI	GI: flexible sigmoidoscopy	DISCOMFORT	Patient
Velmahos GC (2004)[78]	26; PG	RCT	OE	Central venous catheter insertion	COMPLICATIONS: pneumothorax, arterial puncture	Missing
Ahlberg G (2005)[41]	12; PG	RCT	NI	GI: colonoscopy	SUCCESS COMPLICATIONS: no complications DISCOMFORT	Patient
Hochberger J (2005)[42]	28; PG	RCT	NI	GI: upper gastrointestinal endoscopy	SUCCESS COMPLICATIONS: bleeding, esophageal perforation	Patient
Cohen J (2006)[43]	49; PG	RCT	NI	GI: colonoscopy	SUCCESS DISCOMFORT: rated by instructor	Trainee
Thomson M (2006)[44]	14; PG	Obs	NI	GI: colonoscopy	SUCCESS: reach cecum, intubate terminal ileum [tracked but not reported]	Trainee
Ahlberg G (2007)[45]	13; PG	RCT	NI	Surgery: laparoscopic cholecystectomy	COMPLICATIONS: conversion to open procedure	Missing
Davis DP (2007)[85]	120; RN, EMT	Obs	Sim	Airway: intubation	SUCCESS: first attempt, final endotracheal intubation, any airway COMPLICATIONS: hypoxic arrest	Patient
Park J (2007)[46]	28; PG	RCT	NI	GI: colonoscopy	SUCCESS COMPLICATIONS: perforation	Trainee
Chandra DB (2008)[86]	30; O	RCT	Sim	Airway: fiberoptic intubation	SUCCESS	Patient
Draycott TJ (2008)[47]	254; MD,O	1PP	NI	Obstetrics: Shoulder dystocia	COMPLICATIONS: neonatal injury (6 distinct injuries reported)	Patient
Gómez LM (2008)[62]	29; MS	RCT	NI	Airway: intubation	COMPLICATIONS: bradycardia, tachycardia, hypertension, hypoxemia, perioral trauma, laryngospasm, throat pain, other	Trainee
Wayne DB (2008)[48]	78; PG	Obs	NI	CPR: Advanced Cardiac Life Support	SURVIVAL: survive event, survive to discharge D.O.S.: time to discharge or death postevent	Patient
Yi SY (2008)[49]	11; PG	Obs	NI	GI: colonoscopy	SUCCESS DISCOMFORT: abdominal pain, anal discomfort, inflation	Missing
Barsuk JH (2009)[50]	92; PG	Obs	NI	Central venous catheter insertion	COMPLICATIONS: catheter-related bloodstream infection	Patient
Barsuk JH (2009)[51]	41; PG	Obs	NI	Central venous catheter insertion	COMPLICATIONS: arterial puncture, pneumothorax, need for adjustment	Patient

(*continued on next page*)

Table 2. (continued)

Author (year)	Participants N; type	Study design	Comp.	Task	Outcome(s)	Unit of analysis
Barsuk JH (2009)[52]	103; PG	Obs	NI	Central venous catheter insertion	SUCCESS COMPLICATIONS: arterial puncture, need for adjustment, pneumothorax	Patient
Britt RC (2009)[53]	34; PG	RCT	NI	Central venous catheter insertion	SUCCESS COMPLICATIONS: total complications, arterial puncture, pneumothorax, line positioning, bloodstream infection	Patient
Duncan DR (2009)[54]	5; MD	1PP	NI	Thoracentesis	COMPLICATIONS: pneumothorax, chest tube	Patient
Gaies MG (2009)[55]	38; PG	RCT	NI	Venipuncture, peripheral venous cannulation, lumbar puncture	SUCCESS	Patient
Lubin J (2009)[56]	17; RN,EMT	1PP	NI	Airway: intubation	SUCCESS	Missing
Siassakos D (2009)[57]	300; MD,O	1PP	NI	Obstetrics: Umbilical cord prolapse	COMPLICATIONS: ICU admission, low Apgar, fetal bradycardia SURVIVAL: stillbirth	Patient
Sotto JAR (2009)[75]	40; MS	RCT	OE	Peripheral venous cannulation	SUCCESS	Trainee
Andreatta P (2010)[63]	228; PG	1PP	NI	CPR: resuscitation	SURVIVAL	Patient
Capella J (2010)[65]	114; PG,MD,RN	1PP	NI	Trauma management	COMPLICATIONS: (not specified) SURVIVAL D.O.S.: hospital, ICU, ED DELAY: time to FAST scan, time to CT scan OTHER: time to intubation, time to operating room	Patient
Evans LV (2010)[64]	188; PG	RCT	NI	Central venous catheter insertion	SUCCESS: first attempt, overall COMPLICATIONS: pneumothorax, hemothorax, hemomediastinum, vessel laceration, transient dysrhythmia, air embolus, malposition, bloodstream infection	Trainee
Ferlitsch A (2010)[66]	28; PG	RCT	NI	GI: upper gastrointestinal endoscopy	SUCCESS DISCOMFORT: pain, discomfort ACCURACY: major pathological findings	Patient
Haycock A (2010)[76]	40; PG,RN,O	RCT	OE	GI: colonoscopy	SUCCESS	Patient
Kallstrom R (2010)[67]	24; PG	1PP	NI	Transurethral resection prostate	SUCCESS EVAL FINAL: resection weight	Trainee
Nishisaki A (2010)[58]	78; PG	Obs	NI	Airway: intubation	SUCCESS: first attempt, overall success COMPLICATIONS: esophageal intubation, mainstem bronchial intubation, dental/lip trauma	Patient
Smith CC (2010)[59]	52; PG	RCT	NI	Central venous catheter insertion	COMPLICATIONS: pneumothorax, bleeding, arterial puncture	Patient
Tongprasert F (2010)[69]	10; MD	Obs	NI	Obstetrics: Cordocentesis	SUCCESS SURVIVAL: procedure-related loss, total fetal loss	Patient
Weidman EK (2010)[61]	30; PG	RCT	NI	CPR: cardiopulmonary resuscitation	SURVIVAL: survival to discharge, return of spontaneous circulation	Missing
Campos JH (2011)[79]	27; PG,MD	RCT	OE	Airway: intubation	SUCCESS	Trainee
Khouli H (2011)[68]	105; PG	RCT	NI	Central venous catheter insertion	COMPLICATIONS: infection	Patient
Zamora Z (2011)[70]	37; RN	1PP	NI	Bladder irrigation	COMPLICATIONS: need for physician intervention	Patient

Participants: N indicates number of trainees enrolled. (n/a sample size not reported); MS medical student; PG postgraduate physician trainee (resident); MD practicing physician; NS nursing student; RN practicing nurse; EMT emergency medicine technician or other first responder; O other health professional
Study design: 1PP one-group pre-post study; Obs nonrandomized two-group study; RCT randomized two-group study
Comp. (comparison intervention): NI no intervention; OE other (non-simulation) instruction; Sim other simulation
Task: GI gastroenterology; CPR cardiopulmonary resuscitation
Outcome: SUCCESS successful completion; EVAL FINAL evaluation of final result; D.O.S. duration of stay

Table 3. Quality of Included Studies

Scale Item	Subscale (points if present)	No. (%) present
Medical Education Research Study Quality Instrument (MERSQI)*		
Study design (maximum 3)	1-group pre-post (1.5)	9 (18)
	Observational 2-group (2)	14 (28)
	Randomized 2-group (3)	27 (56)
Sampling: No. institutions (maximum 1.5)	1 (0.5)	37 (74)
	2 (1)	4 (8)
	> 2 (1.5)	9 (18)
Sampling: Follow-up (maximum 1.5)	< 50 % or not reported (0.5)	13 (26)
	50–74 % (1)	3 (6)
	≥ 75 % (1.5)	34 (68)
Type of data: Outcome assessment (maximum 3)	Subjective (1)	4 (8)
	Objective (3)	46 (92)
Validity evidence (maximum 3)	Content (1)	7 (14)
	Internal structure (1)	3 (6)
	Relations to other variables (1)	3 (6)
Data analysis: appropriate (maximum 1)	Appropriate (1)	13 (26)
Data analysis: sophistication (maximum 2)	Descriptive (1)	6 (12)
	Beyond descriptive analysis (2)	44 (88)
Highest outcome type (maximum 3)	Patient/health care outcomes (3)	50 (100)
Newcastle-Ottawa Scale (modified)†		
Representativeness of sample	Present (1)	20 (40)
Comparison group from same community	Present (1)	39 (78)
Comparability of comparison cohort, criterion A‡	Present (1)	28 (56)
Comparability of comparison cohort, criterion B‡	Present (1)	17 (34)
Blinded outcome assessment	Present (1)	18 (36)
Follow-up high‡	Present (1)	38 (76)
Other methodological indicators		
Unit of analysis	Appropriate	14 (28)
	Inappropriate	29 (58)
	Not defined	7 (14)
Data linked to trainee	Yes	38 (76)
	No (aggregate, or not reported)	12 (24)
Data source	Trainee	6 (12)
	Patient	3 (6)
	Investigator	25 (50)
	Patient record	16 (320)
Power calculation for patient outcome	Reported	7 (14)
Patient demographic information	Reported	22 (44)
	Adjusted in analysis	5 (10)
	Not reported or adjusted	27 (54)
Patient outcome priority	Primary	13 (26)
	Secondary	5 (10)
	Not reported	32 (64)

N=50

**MERSQI total score (maximum 18): mean 12.5 (SD 1.7), median 12.8 (range 7.5–16.5)*

†*NOS total score (maximum 6): mean 3.2 (SD 1.5), median 3 (range 0–6)*

‡*Comparability of cohorts criterion A was present if the study a) was randomized, or b) controlled for a baseline learning outcome; criterion B was present if a) a randomized study concealed allocation, or b) an observational study controlled for another baseline learner characteristic. Follow-up was high if ≥75 % of those enrolled provided outcome data, or if authors described those lost to follow-up*

interval [CI], 0.31–0.63; $p<.001$), consistent with a small-moderate effect favoring simulation. However, there was large inconsistency among studies, with individual effect sizes ranging from −0.67 to 1.68 and $I^2=69$ %. The funnel plot was symmetric, suggesting that publication bias did not appreciably influence this estimate.

Three studies reported better outcomes from the untrained group, but differences were not statistically significant. A 10-min skill session did not improve first-attempt endotracheal intubation success (20/40 for trained residents vs. 15/24 for untrained).[58] Residents participating in a 10-h simulation-based course had significantly improved adherence to guidelines during cardiac resuscitation, but patient survival to discharge was slightly lower (9/20 for trained vs. 13/28 for untrained).[48] Self-directed hands-on practice with a manikin, added to multimodal training common to all trainees (computer, lecture, observation, and supervised practice on one patient), led to improved instructor ratings of competence, but slightly higher rate of perioral trauma during endotracheal intubation (3/10 for trained vs. 2/13 untrained).[62]

Planned subgroup analyses did not reveal any statistically significant interactions (see eTable 1). Effect sizes were smaller for two-group versus one-group studies (0.40 vs. 0.60, $p_{interaction}=0.25$) and randomized versus nonrandomized studies (0.33 vs. 0.46, $p_{interaction}=0.62$). Effects were larger for studies using the correct versus incorrect unit of analysis (pooled ES 0.53 vs. 0.47, $p_{interaction}=0.85$) and blinded versus unblinded outcome assessment (0.61 vs. 0.41, $p_{interaction}=0.39$). Results for within-patient and to-patient outcomes were virtually identical (pooled ES 0.47 for both).

Sensitivity analysis excluding one study with imprecise effect size calculations did not appreciably alter the results. One group of investigators published a series of studies with overlapping dates, evaluating training in central venous catheterization using outcomes of arterial puncture,[51,52] pneumothorax,[52] procedural success,[52] and catheter-related bloodstream infection.[50] Because of possible non-independence among these three studies, we conducted a sensitivity analysis including only one of these studies, with virtually identical results.

Quantitative Synthesis: Comparative Effectiveness

Nine studies (494 participants) made comparison with non-simulation training (e.g., lecture or standardized patient).[71–79] The pooled effect size (see Fig. 2) for these studies was 0.36 (95 % CI, −0.06 to 0.78; $p=0.09$), consistent with a small effect. Inconsistency was large ($I^2=70$ %) and effect sizes ranged from −1.37 to 1.51. The seven randomized trials had an effect size of 0.53. The funnel plot was symmetric, and sensitivity analyses did not alter conclusions.

Eight studies evaluated the comparative effectiveness of two different simulation-based instructional approaches.[71,80–86] For these studies, we conducted meta-analyses according to seven

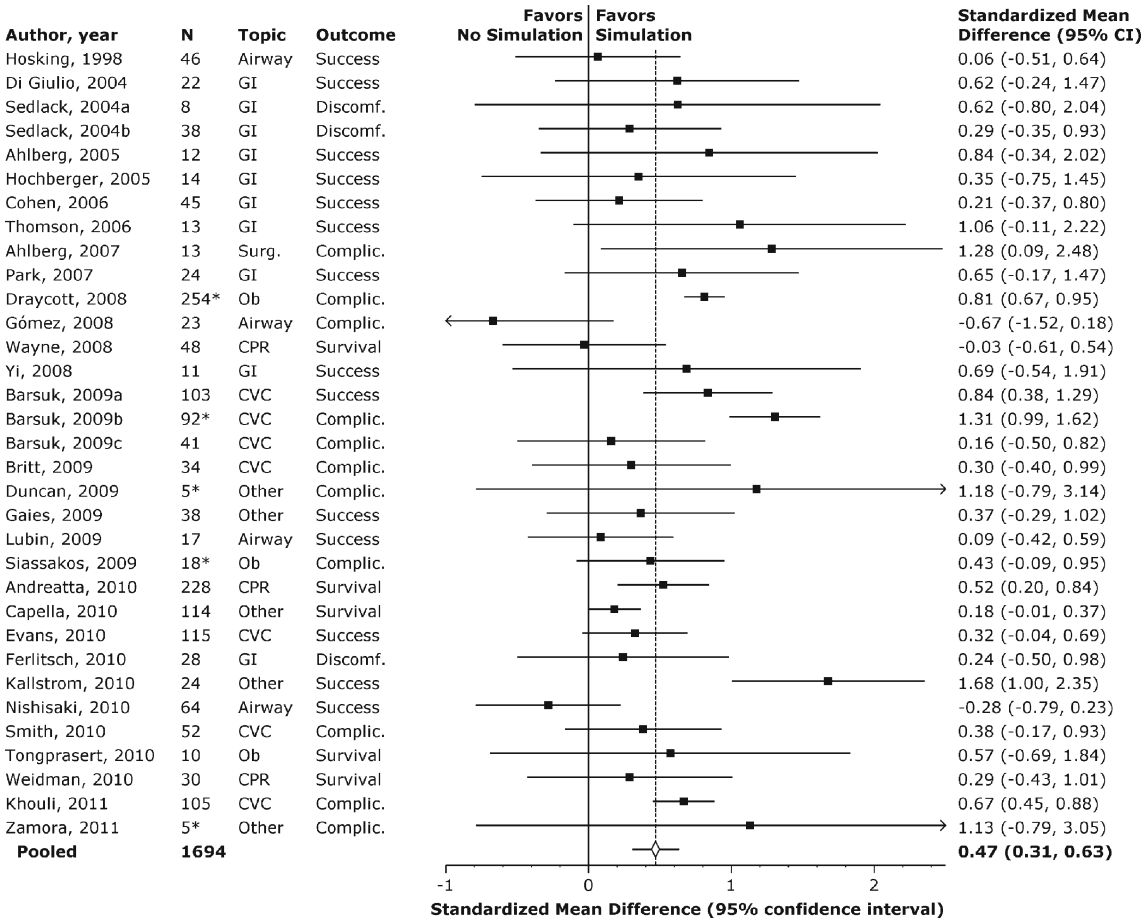

Figure 1. Outcomes of simulation-based education in comparison with no intervention. N indicates number contributing outcomes, except where marked as *, which reflects number enrolled/trained. Abbreviations: *Discomf.* patient discomfort; *Complic.* complications. Clarification of author/year: Sedlack 2004a,[39] Sedlack 2004b,[40] Barsuk 2009a,[52] Barsuk 2009b,[50] Barsuk 2009c[51].

key instructional design features, by looking for differences in the presence of that feature between study arms (see Fig. 3). For example, increased clinical variation among simulated cases has been hypothesized to enhance learning.[34] The degree of clinical variation differed between interventions in three studies, and pooling these results confirmed that more clinical variation was associated with significantly improved outcomes (pooled effect size, 0.46 [95 % CI, 0.18–0.74; p=0.001]). Similar analyses suggested that use of multiple learning strategies and longer time spent learning are associated with improved patient outcomes. For cognitive interactivity, individualized learning, mastery learning, and range of difficulty, the differences were small and not statistically significant.

DISCUSSION

We identified 50 studies reporting patient outcomes in the evaluation of simulation-based education for health professionals. All of these studies involved procedural

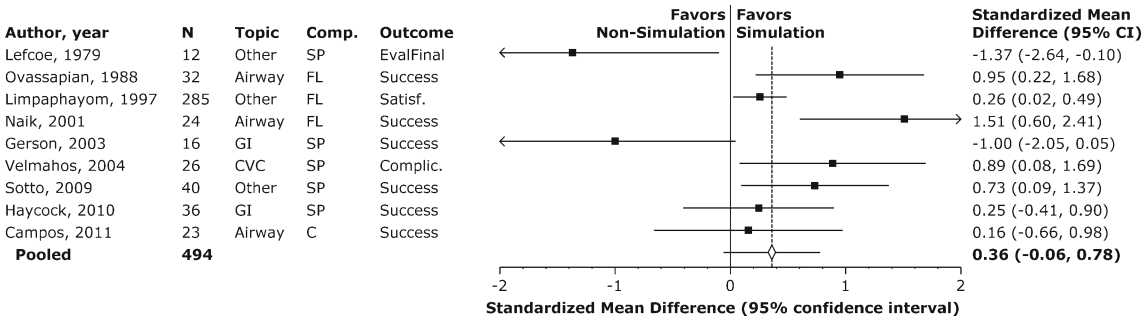

Figure 2. Outcomes of simulation-based education in comparison with non-simulation instruction. N indicates number contributing outcomes. Abbreviations: *Comp.* comparison intervention (*P* standardized or real patient; *FL* face-to-face lecture; *C* computer assisted instruction); *EvalFinal* evaluation of final product; *Satisf.* patient satisfaction; *Complic.* complications.

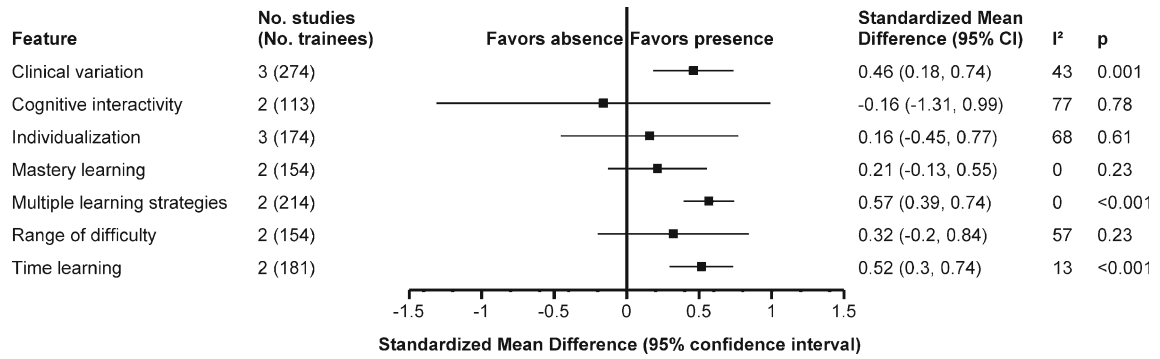

Figure 3. Outcomes of studies comparing two simulation-based educational interventions. N indicates number contributing outcomes.

tasks such as airway management, gastrointestinal endoscopy, or central venous catheter insertion. Most outcomes reflected procedural success or complications, with a small minority reporting other outcomes such as survival and duration of hospitalization.

We found a high prevalence of statistical errors, most notably unit of analysis errors. Few studies assessed outcomes in a blinded fashion, defined a primary outcome, presented measurement validity evidence, or reported sample size estimates. Several studies omitted basic information, such as statistical methods or trainee sample size.

Meta-analytic synthesis demonstrated small-moderate effects favoring simulation in comparison with no intervention, and small nonsignificant effects favoring simulation in comparison with non-simulation instruction, confirming that simulation-based education is associated with downstream benefits on patient care.[4] Instructional design features of clinical variation, more learning strategies, and longer time spent learning were also associated with improved patient outcomes. However, these conclusions are tempered by the fact that only half these studies used trainees as the unit of analysis.

Limitations

We used intentionally broad inclusion criteria, and thus included studies reflecting varied training topics, simulation modalities, and comparison interventions. This increased the number of eligible studies and enhances the generalizability of our findings, but also likely contributed to between-study inconsistencies. To mitigate this limitation, we grouped studies for meta-analysis according to comparison, and explored inconsistency through planned subgroup analyses to investigate possible interactions with topic and study methods.

While we have identified the prevalence of unit of analysis error, we could not determine the degree to which these errors affected study conclusions or meta-analysis results.

Strengths include a novel research question, a comprehensive literature search, rigorous coding with high reproducibility, and the use of meta-analysis to quantitatively synthesize results. We weighted all meta-analyses using the number of trainees.

Integration with Other Literature

Our findings regarding the prevalence of patient outcomes parallel those of previous reviews in medical education.[5–9] To these studies, we add a novel approach to classifying patient outcomes, a careful evaluation of methodological issues, and a quantitative synthesis. Our synthesis of evidence across 50 studies complements and expands upon proposed conceptual models for educationally-relevant patient outcomes.[22, 87, 88]

The reporting of validity evidence in this sample was even less frequent than in previous reviews in medical education.[6, 9, 13, 18–21] We discuss this below. Our findings regarding the prevalence of unit of analysis error mirror those in clinical medicine.[24–29] The overall MERSQI scores of this sample were slightly lower than those of 13 patient outcomes studies of resident shift length.[21]

Implications for Practice and Research

Although this study focused on the field of simulation, we suspect several messages will apply broadly, including the novel within-patient versus to-patient classification framework, the need to avoid unit of analysis error, and the need for evidence to support measurement validity. Regarding the classification framework, early in our review activities we noted a tension between to-patient and within-patient outcomes: outcomes happening to the patient can be determined directly for every procedure, but might not necessarily result in demonstrable morbidity, whereas those arising within the patient are more difficult to measure and likely less sensitive to training. We believe this distinction is important, and that both types are useful and complementary.

Validity evidence was reported much less frequently than in other reviews. We suggest that in most instances it would enhance study rigor to evaluate properties such as interrater reliability for data abstraction or rater observations; content evidence in defining endpoints and complications to monitor; and relations (correlations) between measures such as end-product evaluation, test results, and patient symptoms.

We do not suggest that the pooled effect size for each comparison reflects a single truth applicable across any

simulation-based intervention. Rather, these represent rough estimates of the effect sizes that might be expected. Estimates such as these provide useful information to researchers planning future studies, and may facilitate reversal of the current trend to neglect sample size planning.[89] Researchers should consider the comparison intervention when planning, as pooled effect sizes were generally lower for comparisons with active interventions.

Certain study subgroups were absent or present in lower proportions compared with the larger cohort from which this sample was extracted,[13–15] including nonprocedural activities such as physical exam, patient counseling, or clinical reasoning; procedures such as surgery and anesthesiology; and some learner groups (notably medical students). This suggests selection bias in the topics and learners represented. If, as we suspect, it is more difficult to establish the link between instruction and outcomes for some educational activities than others, then a requirement for patient outcomes in education research could inadvertently exclude important themes from equal status in the literature. Moreover, the benefits of using patient outcomes appear to have come at the price of other methodological weaknesses, since MERSQI total scores are no higher in this sample than in the parent review, even though the MERSQI gives extra weigh to patient outcomes. Researchers, educators, and other stakeholders must consider these and other tradeoffs[90] as they draw inferences about study findings and make decisions based on these inferences.

Acknowledgements:
Contributors: *The authors thank Rose Hatala, MD, MSc, Stanley J. Hamstra, Phd, Jason H. Szostek, MD, and Patricia J. Erwin, MLS, for their assistance in the literature search and initial data acquisition*

Funders: *This work was supported by intramural funds, including an award from the Division of General Internal Medicine, Mayo Clinic. The funding sources for this study played no role in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation of the manuscript. The funding sources did not review the manuscript.*

Conflicts of Interest: *No relevant conflicts of interest.*

Corresponding Author: *David A. Cook, MD, MHPE; Division of General Internal Medicine, Mayo Clinic College of Medicine, Mayo 17, 200 First Street SW, Rochester, MN 55905, USA (e-mail: cook.david33@mayo.edu).*

REFERENCES

1. **Chen FM, Bauchner H, Burstin H.** A call for outcomes research in medical education. Acad Med. 2004;79:955–60.
2. **Dauphinee WD.** Educators must consider patient outcomes when assessing the impact of clinical training. Med Educ. 2012;46:13–20.
3. **Shea JA.** Mind the gap: some reasons why medical education research is different from health services research. Med Educ. 2001;35:319–20.
4. **McGaghie WC.** Medical education research as translational science. Sci Transl Med. 2010;2:19cm18.
5. **Prystowsky JB, Bordage G.** An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. Med Educ. 2001;35:331–6.
6. **Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM.** Association between funding and quality of published medical education research. JAMA. 2007;298:1002–9.
7. **Baernstein A, Liss HK, Carney PA, Elmore JG.** Trends in study methods used in undergraduate medical education research, 1969–2007. JAMA. 2007;298:1038–45.
8. **Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA.** Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's medical education special issue. J Gen Intern Med. 2008;23:903–7.
9. **Cook DA, Levinson AJ, Garside S.** Method and reporting quality in health professions education research: a systematic review. Med Educ. 2011;45:227–38.
10. **Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, et al.** Simulation technology for health care professional skills training and assessment. JAMA. 1999;282:861–6.
11. **Gaba DM.** The future vision of simulation in healthcare. Simul Healthc. 2007;2:126–35.
12. **McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ.** A critical review of simulation-based medical education research: 2003–2009. Med Educ. 2010;44:50–63.
13. **Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al.** Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA. 2011;306:978–88.
14. **Cook DA, Brydges R, Hamstra S, Zendejas B, Szostek JH, Wang AT, et al.** Comparative Effectiveness of Technology-Enhanced Simulation vs. Other Instructional Methods: A Systematic Review and Meta-Analysis. *Simul Healthc.* 2012;7:308–20.
15. **Cook DA, Hamstra S, Brydges R, Zendejas B, Szostek JH, Wang AT, et al.** Comparative Effectiveness of Instructional Design Features in Simulation-based Education: Systematic Review and Meta-analysis. *Med Teach.* 2012; Online early (doi:10.3109/0142159X.2012.714886).
16. **Downing SM.** Validity: on the meaningful interpretation of assessment data. Med Educ. 2003;37:830–7.
17. **Cook DA, Beckman TJ.** Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119:166.e7–16.
18. **Beckman TJ, Cook DA, Mandrekar JN.** What is the validity evidence for assessments of clinical teaching? J Gen Intern Med. 2005;20:1159–64.
19. **Kogan JR, Holmboe ES, Hauer KE.** Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. JAMA. 2009;302:1316–26.
20. **Ratanawongsa N, Thomas PA, Marinopoulos SS, Dorman T, Wilson LM, Ashar BH, et al.** The reported validity and reliability of methods for evaluating continuing medical education: a systematic review. Acad Med. 2008;83:274–83.
21. **Reed DA, Fletcher KE, Arora VM.** Systematic review: association of shift length, protected sleep time, and night float with patient care, residents' health, and education. Ann Intern Med. 2010;153:829–42.
22. **Kalet AL, Gillespie CC, Schwartz MD, Holmboe ES, Ark TK, Jay M, et al.** New measures to establish the evidence base for medical education: identifying educationally sensitive patient outcomes. Acad Med. 2010;85:844–51.
23. **Bland JM, Kerry SM.** Trials randomised in clusters (statistics notes). BMJ. 1997;315:600.
24. **Donner A.** An empirical study of cluster randomization. Int J Epidemiol. 1982;11:283–6.
25. **Whiting-O'Keefe QE, Henke C, Simborg DW.** Choosing the correct unit of analysis in medical care experiments. Med Care. 1984;22:1101–14.
26. **Divine GW, Brown JT, Frazier LM.** The unit of analysis error in studies about physicians' patient care behavior. J Gen Intern Med. 1992;7: 623–9.
27. **Thomas RE, Ramsay CR, McAuley L, Grimshaw JM.** Unit of analysis errors should be clarified in meta-analyses. BMJ. 2003;326:397.
28. **Calhoun AW, Guyatt GH, Cabana MD, Lu D, Turner DA, Valentine S, et al.** Addressing the unit of analysis in medical care studies: a systematic review. Med Care. 2008;46:635–43.
29. **Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G.** Internal and external validity of cluster randomised trials: systematic review of recent trials. BMJ. 2008;336:876–80.
30. **Kerry SM, Bland JM.** Analysis of a trial randomised in clusters (Statistics notes). BMJ. 1998;316:54.

31. **Campbell MK, Elbourne DR, Altman DG.** CONSORT statement: extension to cluster randomised trials. BMJ. 2004;328:702–8.
32. **Moher D, Liberati A, Tetzlaff J, Altman DG.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151:264–9.
33. **Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM.** Internet-based learning in the health professions: a meta-analysis. JAMA. 2008;300:1181–96.
34. **Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ.** Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;27:10–28.
35. **Higgins JPT, Thompson SG, Deeks JJ, Altman DG.** Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.
36. **Cohen J.** Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
37. **Swanson ER, Fosnocht DE.** Effect of an airway education program on prehospital intubation. Air Med J. 2002;21:28–31.
38. **Di Giulio E, Fregonese D, Casetti T, Cestari R, Chilovi F, D'Ambra G, et al.** Training with a computer-based simulator achieves basic manual skills required for upper endoscopy: a randomized controlled trial. Gastrointest Endosc. 2004;60:196–200.
39. **Sedlack RE, Kolars JC.** Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. Am J Gastroenterol. 2004;99:33–7.
40. **Sedlack RE, Kolars JC, Alexander JA.** Computer simulation training enhances patient comfort during endoscopy. Clin Gastroenterol Hepatol. 2004;2:348–52.
41. **Ahlberg G, Hultcrantz R, Jaramillo E, Lindblom A, Arvidsson D.** Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? Endoscopy. 2005;37:1198–204.
42. **Hochberger J, Matthes K, Maiss J, Koebnick C, Hahn EG, Cohen J.** Training with the compactEASIE biologic endoscopy simulator significantly improves hemostatic technical skill of gastroenterology fellows: a randomized controlled comparison with clinical endoscopy training alone. Gastrointest Endosc. 2005;61:204–15.
43. **Cohen J, Cohen SA, Vora KC, Xue X, Burdick JS, Bank S, et al.** Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. Gastrointest Endosc. 2006;64:361–8.
44. **Thomson M, Heuschkel R, Donaldson N, Murch S, Hinds R.** Acquisition of competence in paediatric ileocolonoscopy with virtual endoscopy training. J Pediatr Gastroenterol Nutr. 2006;43:699–701.
45. **Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA III, et al.** Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg. 2007;193:797–804.
46. **Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, et al.** Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg. 2007;194:205–11.
47. **Draycott TJ, Crofts JF, Ash JP, Wilson LV, Yard E, Sibanda T, et al.** Improving neonatal outcome through practical shoulder dystocia training. Obstet Gynecol. 2008;112:14–20.
48. **Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie WC.** Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. Chest. 2008;133:56–61.
49. **Yi SY, Ryu KH, Na YJ, Woo HS, Ahn W, Kim WS, et al.** Improvement of colonoscopy skills through simulation-based training. Stud Health Technol Inform. 2008;132:565–7.
50. **Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB.** Use of simulation-based education to reduce catheter-related bloodstream infections. Arch Intern Med. 2009;169:1420–3.
51. **Barsuk JH, McGaghie WC, Cohen ER, Balachandran JS, Wayne DB.** Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit. J Hosp Med. 2009;4:397–403.
52. **Barsuk JH, McGaghie WC, Cohen ER, O'Leary KJ, Wayne DB.** Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. Crit Care Med. 2009;37:2697–701.
53. **Britt RC, Novosel TJ, Britt LD, Sullivan M.** The impact of central line simulation before the ICU experience. Am J Surg. 2009;197:533–6.
54. **Duncan DR, Morgenthaler TI, Ryu JH, Daniels CE.** Reducing iatrogenic risk in thoracentesis: establishing best practice via experiential training in a zero-risk environment. Chest. 2009;135:1315–20.
55. **Gaies MG, Morris SA, Hafler JP, Graham DA, Capraro AJ, Zhou J, et al.** Reforming procedural skills training for pediatric residents: a randomized, interventional trial. Pediatrics. 2009;124:610–9.
56. **Lubin J, Carter R.** The feasibility of daily mannequin practice to improve intubation success. Air Med J. 2009;28:195–7.
57. **Siassakos D, Hasafa Z, Sibanda T, Fox R, Donald F, Winter C, et al.** Retrospective cohort study of diagnosis—delivery interval with umbilical cord prolapse: the effect of team training. BJOG. 2009;116:1089–96.
58. **Nishisaki A, Donoghue AJ, Colborn S, Watson C, Meyer A, Brown CA III, et al.** Effect of just-in-time simulation training on tracheal intubation procedure safety in the pediatric intensive care unit. Anesthesiology. 2010;113:214–23.
59. **Smith CC, Huang GC, Newman LR, Clardy PF, Feller-Kopman D, Cho M, et al.** Simulation training and its effect on long-term resident performance in central venous catheterization. Simul Healthc. 2010;5:146–51.
60. **Hosking EJ.** Does practising intubation on a manikin improve both understanding and clinical performance of the task by medical students? Anaesthesia Points West. 1998;31:25–8.
61. **Weidman EK, Bell G, Walsh D, Small S, Edelson DP.** Assessing the impact of immersive simulation on clinical performance during actual in-hospital cardiac arrest with CPR-sensing technology: a randomized feasibility study. Resuscitation. 2010;81:1556–61.
62. **Gómez LM, Calderón M, Sáenz X, Reyes G, Moreno MA, Ramírez LJ, et al.** Effect and benefit of clinical simulation in the development of psychomotor competencies in anesthesia: a random double-blind clinical trial [Spanish]. Revista Colombiana de Anestesiologia. 2008;36:93–107.
63. **Andreatta P, Saxton E, Thompson M, Annich G.** Simulation-based mock codes significantly correlate with improved pediatric patient cardiopulmonary arrest survival rates. Pediatr Crit Care Med. 2010;12:33–8.
64. **Evans LV, Dodge KL, Shah TD, Kaplan LJ, Siegel MD, Moore CL, et al.** Simulation training in central venous catheter insertion: improved performance in clinical practice. Acad Med. 2010;85:1462–9.
65. **Capella J, Smith S, Philp A, Putnam T, Gilbert C, Fry W, et al.** Teamwork training improves the clinical care of trauma patients. J Surg Educ. 2010;67:439–43.
66. **Ferlitsch A, Schoefl R, Puespoek A, Miehsler W, Schoeniger-Hekele M, Hofer H, et al.** Effect of virtual endoscopy simulator training on performance of upper gastrointestinal endoscopy in patients: a randomized controlled trial. Endoscopy. 2010;42:1049–56.
67. **Kallstrom R, Hjertberg H, Svanvik J.** Impact of virtual reality-simulated training on urology residents' performance of transurethral resection of the prostate. J Endourol. 2010;24:1521–8.
68. **Khouli H, Jahnes K, Shapiro J, Rose K, Mathew J, Gohil A, et al.** Performance of medical residents in sterile techniques during central vein catheterization: randomized trial of efficacy of simulation-based training. Chest. 2011;139:80–7.
69. **Tongprasert F, Wanapirak C, Sirichotiyakul S, Piyamongkol W, Tongsong T.** Training in cordocentesis: the first 50 case experience with and without a cordocentesis training model. Prenat Diagn. 2010;30:467–70.
70. **Zamora Z, Clark MJ, Winslow B, Schatzschneider M, Burkard J.** Orthotopic neobladder irrigation: competency assessment through simulation. Urol Nurs. 2011;31:113–20.
71. **Lefcoe DL, Green ML.** Simulated models: a mode for instruction in root planing procedures. Educ Dir Dent Aux. 1979;3:20–4.
72. **Ovassapian A, Yelich SJ, Dykes MH, Golman ME.** Learning fibreoptic intubation: use of simulators v. traditional teaching. Br J Anaesth. 1988;61:217–20.
73. **Limpaphayom K, Ajello C, Reinprayoon D, Lumbiganon P, Graffikin L.** The effectiveness of model-based training in accelerating IUD skill acquisition. A study of midwives in Thailand. Br J Fam Plann. 1997;23:58–61.
74. **Gerson LB, Van Dam J.** A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. Endoscopy. 2003;35:569–75.
75. **Sotto JAR, Ayuste EC Jr, Bowyer MW, Almonte JR, Dofitas RB, Lapitan MCM, et al.** Exporting simulation technology to the Philippines: a comparative study of traditional versus simulation methods for teaching intravenous cannulation. Stud Health Technol Inform. 2009;142:346–51.
76. **Haycock A, Koch AD, Familiari P, van Delft F, Dekker E, Petruzziello L, et al.** Training and transfer of colonoscopy skills: a multinational, randomized, blinded, controlled trial of simulator versus bedside training. Gastrointest Endosc. 2010;71:298–307.

77. **Naik VN, Matsumoto ED, Houston PL, Hamstra SJ, Yeung RY-M, Mallon JS, et al.** Fiberoptic orotracheal intubation on anesthetized patients: do manipulation skills learned on a simple model transfer into the operating room? Anesthesiology. 2001;95: 343–8.
78. **Velmahos GC, Toutouzas KG, Sillin LF, Chan L, Clark RE, Theodorou D, et al.** Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory. Am J Surg. 2004;187:114–9.
79. **Campos JH, Hallam EA, Ueda K.** Training in placement of the left-sided double-lumen tube among non-thoracic anaesthesiologists: intubation model simulator versus computer-based digital video disc, a randomised controlled trial. Eur J Anaesthesiol. 2011;28: 169–74.
80. **Stewart RD, Paris PM, Pelton GH, Garretson D.** Effect of varied training techniques on field endotracheal intubation success rates. Ann Emerg Med. 1984;13:1032–6.
81. **Stratton SJ, Kane G, Gunter CS, Wheeler NC, Ableson-Ward C, Reich E, et al.** Prospective study of manikin-only versus manikin and human subject endotracheal intubation training of paramedics. Ann Emerg Med. 1991;20:1314–8.
82. **Trooskin SZ, Rabinowitz S, Eldridge C, McGowan DE, Flancbaum L.** Teaching endotracheal intubation using animals and cadavers. Prehospital Disaster Med. 1992;7:179–82.
83. **Chang KK, Chung JW, Wong TK.** Learning intravenous cannulation: a comparison of the conventional method and the CathSim intravenous training system. J Clin Nurs. 2002;11:73–8.
84. **Rumball C, Macdonald D, Barber P, Wong H, Smecher C.** Endotracheal intubation and esophageal tracheal Combitube insertion by regular ambulance attendants: a comparative trial. Prehosp Emerg Care. 2004;8:15–22.
85. **Davis DP, Buono C, Ford J, Paulson L, Koenig W, Carrison D.** The effectiveness of a novel, algorithm-based difficult airway curriculum for air medical crews using human patient simulators. Prehosp Emerg Care. 2007;11:72–9.
86. **Chandra DB, Savoldelli GL, Joo HS, Weiss ID, Naik VN.** Fiberoptic oral intubation: the effect of model fidelity on training for transfer to patient care. Anesthesiology. 2008;109:1007–13.
87. **Haan CK, Edwards FH, Poole B, Godley M, Genuardi FJ, Zenni EA.** A model to begin to use clinical outcomes in medical education. Acad Med. 2008;83:574–80.
88. **Swing SR, Schneider S, Bizovi K, Chapman D, Graff LG, Hobgood C, et al.** Using patient care quality measures to assess educational outcomes. Acad Emerg Med. 2007;14:463–73.
89. **Cook DA.** If you teach them, they will learn: why medical education needs comparative effectiveness research. Adv Health Sci Educ Theory Pract. 2012;17:305–10.
90. **Cook DA, West CP.** Reconsidering the focus on "outcomes research" in medical education: a cautionary note. *Acad Med.* 2013; Accepted (in press).