



## Simulation-Based Bronchoscopy Training Systematic Review and Meta-analysis

Cassie C. Kennedy, MD, FCCP; Fabien Maldonado, MD, FCCP; and David A. Cook, MD

**Background:** Simulation-based bronchoscopy training is increasingly used, but effectiveness remains uncertain. We sought to perform a comprehensive synthesis of published work on simulation-based bronchoscopy training.

**Methods:** We searched MEDLINE, EMBASE, CINAHL, PsycINFO, ERIC, Web of Science, and Scopus for eligible articles through May 11, 2011. We included all original studies involving health professionals that evaluated, in comparison with no intervention or an alternative instructional approach, simulation-based training for flexible or rigid bronchoscopy. Study selection and data abstraction were performed independently and in duplicate. We pooled results using random effects meta-analysis.

**Results:** From an initial pool of 10,903 articles, we identified 17 studies evaluating simulation-based bronchoscopy training. In comparison with no intervention, simulation training was associated with large benefits on skills and behaviors (pooled effect size, 1.21 [95% CI, 0.82-1.60]; n = 8 studies) and moderate benefits on time (0.62 [95% CI, 0.12-1.13]; n = 7). In comparison with clinical instruction, behaviors with real patients showed nonsignificant effects favoring simulation for time (0.61 [95% CI, -1.47 to 2.69]) and process (0.33 [95% CI, -1.46 to 2.11]) outcomes (n = 2 studies each), although variation in training time might account for these differences. Four studies compared alternate simulation-based training approaches. Inductive analysis to inform instructional design suggested that longer or more structured training is more effective, authentic clinical context adds value, and animal models and plastic part-task models may be superior to more costly virtual-reality simulators.

**Conclusions:** Simulation-based bronchoscopy training is effective in comparison with no intervention. Comparative effectiveness studies are few. *CHEST* 2013; 144(1):183-192

Heightened concerns for patient safety have prompted an ongoing shift from apprenticeship models of medical education to approaches that insulate patients from the initial learning phase in procedural training.<sup>1</sup>

Manuscript received July 20, 2012; revision accepted January 2, 2013.

**Affiliations:** From the Division of Pulmonary and Critical Care Medicine (Drs Kennedy and Maldonado) and Division of General Internal Medicine (Dr Cook), Mayo Clinic; and Office of Education Research (Dr Cook), Mayo Medical School, Mayo Foundation for Medical Education and Research, Rochester, MN.

**Funding/Support:** This work was supported by intramural funds, including an award from the Division of General Internal Medicine, Mayo Clinic.

**Correspondence to:** David A. Cook, MD, Division of General Internal Medicine, College of Medicine, Mayo Clinic, Mayo 17-W, 200 First St SW, Rochester, MN 55905; e-mail: cook.david33@mayo.edu

© 2013 American College of Chest Physicians. Reproduction of this article is prohibited without written permission from the American College of Chest Physicians. See online for more details. DOI: 10.1378/chest.12-1786

Simulation-based education appears ideally suited to offer effective training in a zero-risk environment.<sup>2</sup> Simulation has been widely embraced by the critical care community, as it decreases patient risks and enables deliberate structuring of instruction including exposure to rare events and diseases.<sup>3</sup> Simulation-based training has shown efficacy for tasks such as endotracheal intubation,<sup>4,5</sup> central venous catheterization,<sup>6</sup> and cardiopulmonary resuscitation.<sup>7</sup> Yet training for bronchoscopy, a common procedure for pulmonologists, varies substantially across teaching programs.<sup>8,9</sup> A synthesis of evidence regarding the effectiveness and key features of simulation-based bronchoscopy training would facilitate decision-making among educators and highlight remaining research needs.

Two previous reviews of bronchoscopy training have offered such summaries. One systematic review provided a narrative summary of seven comparative studies<sup>10</sup>; the other presented a nonsystematic review

of available technologies.<sup>11</sup> These reviews were limited by incomplete assessment of study quality and absence of quantitative pooling to derive best estimates of the effect of these interventions on educational outcomes. Moreover, several new studies have been published since these reviews were written, and, thus, an updated synthesis of the evidence would be in order. We sought to identify and quantitatively summarize all comparative studies of simulation-based bronchoscopy training by conducting a systematic review of the literature.

## MATERIALS AND METHODS

This report is a subanalysis of data collected as part of a comprehensive review of simulation-based education<sup>12</sup> that was planned, conducted, and reported in adherence to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards of quality for reporting meta-analyses.<sup>13</sup> Our general methods have been described in detail previously<sup>12</sup>; we summarize them briefly below.

### Questions

We sought to answer the following questions: What is the effectiveness of technology-enhanced simulation for training in bronchoscopy, and what instructional design features are associated with improved outcomes? We defined technology-enhanced simulation as an educational tool or device with which the learner physically interacts to mimic an aspect of clinical care.

### Study Eligibility

We included studies involving health-profession learners at any stage in training or practice that investigated the use of technology-enhanced simulation to teach bronchoscopy in comparison with no intervention (ie, a control arm or preintervention assessment) or an active simulation-based or nonsimulation training activity. We included single-group pretest-posttest and two-group randomized and nonrandomized studies. We included training for flexible bronchoscopy, interventional bronchoscopy including lymph node sampling, rigid bronchoscopy, and foreign body removal. We did not make exclusions based on outcome, or year or language of publication.

### Study Identification

With the assistance of an experienced research librarian, we searched MEDLINE, EMBASE, CINAHL, PsycINFO, ERIC, Web of Science, and Scopus for eligible articles. Our full search strategy has been published previously.<sup>12</sup> The last full search update was May 11, 2011. To identify more recently published articles, we searched PubMed on January 31, 2012, with the key words: bronchosc\* AND (simulat\* OR curric\* or teach\*). We searched for omitted articles by reviewing the reference lists from all included articles, the two reviews of bronchoscopy simulation,<sup>10,11</sup> and several published reviews of health-profession simulation. Finally, we searched the full table of contents of two journals devoted to health-profession simulation (*Simulation in Healthcare* and *Clinical Simulation in Nursing*) and one journal devoted to bronchoscopy (*Journal of Bronchology & Interventional Pulmonology*).

### Study Selection

Study selection involved two stages. In stage 1, we identified all studies of technology-enhanced simulation for health-profession

education. Two reviewers, working independently, screened all titles and abstracts for inclusion. We obtained the full text of all articles that could not be confidently excluded and reviewed these for definitive inclusion/exclusion, again independently and in duplicate. We resolved conflicts by consensus. Chance-adjusted interrater agreement for study inclusion was substantial (intraclass correlation coefficient, 0.69).<sup>14</sup> In stage 2, from this set of articles, we identified studies focused on bronchoscopy training by coding the clinical topic during data extraction.

### Data Extraction

Foreign-language articles were translated prior to data abstraction. We abstracted information independently and in duplicate for all variables where reviewer judgment was required, and resolved conflicts by consensus. We used a data abstraction form to abstract information on the training level of learners, clinical topic, method of group assignment, outcomes, and methodologic quality. Methodologic quality was graded using the Medical Education Research Study Quality Instrument<sup>15</sup> and an adaptation of the Newcastle-Ottawa scale for cohort studies.<sup>16</sup> We also coded features of the intervention, including feedback, repetitions, clinical variation, mastery learning, and whether training took place in a single day or was distributed over > 1 day.

We abstracted information separately for learning outcomes of satisfaction, knowledge, skills, behaviors with patients, and patient effects.<sup>17</sup> We further classified skill as time (time to complete the task) and process (observed proficiency, economy of movements, or minor errors). We also classified behaviors with patients as time and process. We converted reported results to a standardized mean difference (Hedges' *g* effect size) using methods described previously.<sup>12</sup> For articles containing insufficient information to calculate an effect size, we requested information from authors.

### Data Synthesis

We grouped studies according to the comparison arm (comparison with no intervention, nonsimulation intervention, or simulation intervention). For quantitative synthesis, we planned to pool, using meta-analysis, the results of studies comparing simulation training with no intervention. We also planned meta-analysis of studies making comparison with another active instructional intervention whenever two or more studies made a sufficiently similar comparison. We planned sensitivity analyses excluding nonrandomized studies and studies with imprecise effect size calculations (calculated using *P* value upper limits or imputed standard deviations).

We quantified the inconsistency (heterogeneity) across studies using the *I*<sup>2</sup> statistic,<sup>18</sup> which estimates the percentage of variability across studies not due to chance. *I*<sup>2</sup> values > 50% indicate large inconsistency. Because we found large inconsistency in most analyses, we used random effects models to pool weighted effect sizes. We used SAS 9.1 (SAS Institute, Inc) for all analyses. Statistical significance was defined by a two-sided  $\alpha$  of 0.05, and interpretations of clinical significance emphasized CIs in relation to Cohen's effect size classifications (> 0.8 = large, 0.5-0.8 = moderate, and 0.2-0.5 = small).<sup>19</sup> We evaluated the possibility of publication bias using funnel plots in combination with the Eggers asymmetry test, recognizing that these procedures can be misleading in the presence of inconsistency.

We also conducted a nonquantitative synthesis of evidence focused on key instructional design features. We reviewed all comparative-effectiveness studies (studies making comparison with another active instructional intervention) to inductively identify salient themes. We used as an initial guide the 10 key features of high-fidelity simulation identified by Issenberg,<sup>20</sup> but also remained open to the possible presence of other design features.

## RESULTS

### Trial Flow

Using our broad search strategy, we identified 10,903 articles from which we identified 985 comparative studies of technology-enhanced simulation for health-profession training (Fig 1). From this set, we identified 16 articles focused on bronchoscopy training. We did not identify any additional articles in our review of reference lists of included articles or the previously published reviews. However, we identified one additional article in our updated PubMed search, yielding a total of 17 studies eligible for inclusion.<sup>21-37</sup> Nine of these studies made comparison with no intervention,<sup>21,23,25,26,28,31,33-35</sup> four made comparison with nonsimulation training,<sup>22,29,36,37</sup> and four made comparison with another simulation-based instructional approach.<sup>24,27,30,32</sup> One article was published in Swedish<sup>24</sup> and one in Chinese.<sup>31</sup>

### Study Characteristics

Study characteristics are detailed in Table 1 and summarized in Table 2. Learners were most often postgraduate trainees with little or no bronchoscopy experience. Most studies (n = 16) included training for basic flexible bronchoscopy skills (ie, airway inspection). A minority of studies addressed more advanced techniques such as endobronchial ultrasound (two studies), transbronchial needle aspiration (two studies), and rigid bronchoscopy with foreign-body extraction

(two studies). Two studies included practice on pediatric airways.<sup>30,35</sup>

Simulation modalities varied widely. In 12 studies, the simulation intervention comprised a virtual-reality bronchoscopy simulator. Such simulators typically consist of a proxy flexible bronchoscope introduced into an interface device that transmits movements to a computer. The computer displays images of the upper and lower airways on a screen. Many systems provide realistic force feedback and reproduce events such as patient breathing, coughing, and bleeding. Virtual-reality simulators range in price from \$25,000 to > \$100,000. Six studies used part-task models made of synthetics (plastic, silicone, or rubber) or wood. Synthetic models reproduce branching channels that mimic the human bronchial tree. The wood “choose-the-hole” box comprises three to five wood panels drilled with holes through which the endoscope is sequentially navigated. Two studies used mannequins, and two used cadaveric animal tissue.

Regarding instructional design, six interventions distributed training over > 1 day,<sup>21,22,27,33,36,37</sup> and four required trainees to complete > 10 cases.<sup>22,33,36,37</sup> The instructors planned case-to-case clinical variation in six studies.<sup>21,22,26,29,36,37</sup> Only one study provided high (intense or multimodal) performance feedback.<sup>22</sup>

The outcomes studied included computer-generated estimates of performance, subjective or objective assessments by observers, and satisfaction questionnaires. Four studies assessed performance in the context of bronchoscopy of real patients.<sup>22,26,33,37</sup> All of these studies included outcomes focused on the process of bronchoscopy (eg, airway inspection technique or completeness), three used procedural time,<sup>22,26,37</sup> and one used procedural success (ie, successful needle aspiration).<sup>37</sup> Most of the remaining studies used simulation-based outcome measures including combinations of dexterity, accuracy, and speed (economy of performance). Three studies assessed reaction/satisfaction of trainees after simulation training, and two assessed knowledge.

### Study Quality

Study quality is summarized in Table 3. Six studies used a single-group pretest-posttest design. Of the 11 two-group studies, six used randomized group allocation. Five studies lost > 25% of enrolled participants prior to follow-up or failed to report the number of participants included in analysis. Validity evidence to support outcome assessments was reported infrequently: Three studies reported content evidence,<sup>27,33,34</sup> three studies reported relations with other variables,<sup>25,27,33</sup> and only one study reported score reliability.<sup>33</sup>

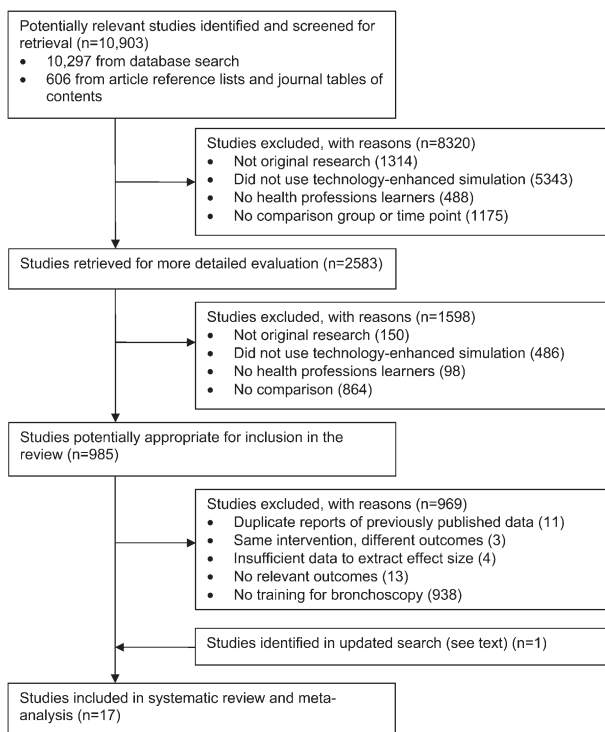


FIGURE 1. Trial flow.

Table 1—List of Included Studies and Key Features

Study Year	Trainee No., Level <sup>a</sup>	Study Design <sup>b</sup>	Comparison <sup>c</sup>	Task <sup>d</sup>	Modality <sup>e</sup>	Instructional Features <sup>f</sup>	Outcomes Reported <sup>g</sup>	Follow-up <sup>h</sup>	Outcome Blind <sup>i</sup>	Outcome Objective <sup>j</sup>	MERSQI Total <sup>k</sup>	NOS Total <sup>l</sup>
Colt <sup>21</sup> /2001	5; PG	1PP	NI	Flex	VR	CV,DP	ST,SP	High	Blind	Obj	11	2
Os <sup>22</sup> /2001	6; PG	RCT	OE	Flex	VR	CV,DP,FB,R	BT,BP	High	Blind	Obj	13	4
Hilmi <sup>23</sup> /2002	20; PG	1PP	NI	Rigid	Animal	M	ST,SP	...	...	Obj	10	0
Ameur <sup>24</sup> /2003	20; MS	2NR	SS	Flex	VR	...	R,SP	...	Blind	Obj	10.5	1
Moorthy <sup>25</sup> /2003	9; O	1PP	NI	Flex	VR	...	ST,SP	...	Blind	Obj	11	1
Blum <sup>26</sup> /2004	10; PG	RCT	NI	Flex	VR	CV	BT,BP	High	Blind	Obj	13	4
Martin <sup>27</sup> /2004	43; PG,MD,O	RCT	SS	Flex	Model	DP	ST,SP	High	Blind	Obj	14.5	5
Agre <sup>28</sup> /2005	5; PG	1PP	NI	Flex	Model	M	ST,SP	...	...	Obj	9	0
Chen <sup>29</sup> /2006	20; O	RCT	OE	Flex	VR	CV	R	High	...	Subj	10	3
Deutsch <sup>30</sup> /2009	36; PG	2NR	SS	Flex, Peds	VR, Mannequin, Animal	...	SP	...	...	Subj	9.5	1
Bian <sup>31</sup> /2010	40; PG	RCT	NI	Flex	Model	...	ST	High	Blind	Obj	12.5	4
Davoudi <sup>32</sup> /2010	48; MD	RCT	SS	Flex, TBNA	VR, Model	...	R	High	...	Subj	9	3
Wahid <sup>33</sup> /2010	47; PG	2NR	NI	Flex	VR	DPR	K,BP	High	...	Obj	16	2
Colt <sup>34</sup> /2011	42; PG	1PP	NI	Flex	Model	...	K,SP	High	Blind	Obj	13	3
Jabhour <sup>35</sup> /2011	17; PG	1PP	NI	Flex, Rigid, Peds	VR, Model, Mannequin	...	ST,SP	High	Blind	Obj	10	2
Stather <sup>36</sup> /2011	13; PG	2NR	OE	Flex, US	VR	CV,DPR	ST,SP	High	Blind	Obj	10.5	3
Stather <sup>37</sup> /2012	8; PG	2NR	OE	Flex, US, TBNA	VR	CV,DPR	BT,BP,P	High	...	Obj	12	2

<sup>a</sup>MD = practicing physician; MS = medical student; O = other health professional; PG = postgraduate physician trainee (resident).

<sup>b</sup>1PP = 1 group pre-post study; 2NR = nonrandomized 2-group study; RCT = randomized 2-group study.

<sup>c</sup>NI = no intervention; OE = other (nonsimulation) education; SS = simulation-simulation comparison.

<sup>d</sup>Flex = flexible bronchoscopy; Peds = pediatric patients; Rigid = rigid bronchoscopy; TBNA = transbronchial needle aspiration; US = endobronchial ultrasound.

<sup>e</sup>VR = virtual reality.

<sup>f</sup>CV = clinical variation present; DP = distributed practice (over > 1 d); FB = feedback high; M = mastery learning model present; R = many repetitions present.

<sup>g</sup>BP = process behavior; BT = time behavior; K = knowledge; P = patient effects; R = reaction (satisfaction); SP = process skill; ST = time skill.

<sup>h</sup>High =  $\geq 75\%$  follow-up for any outcome.

<sup>i</sup>Blinded assessment of highest-level outcome.

<sup>j</sup>Obj = objective assessment; Subj = self-report of highest-level outcome.

<sup>k</sup>MERSQI = Medical Education Research Study Quality Instrument (maximum score 18).

<sup>l</sup>NOS = modified Newcastle-Ottawa scale (maximum score 6).

**Table 2—Summary of Key Features of Included Studies**

Study Characteristic	Level	No. of Studies (No. of Participants <sup>a</sup> )	Relevant References
All studies	...	17 (389)	...
Study design	Posttest-only two-group	11 (291)	22, 24, 26, 27, 29-33, 36, 37
	Pretest-posttest one-group	6 (98)	21, 23, 25, 28, 34, 35
Group allocation	Randomized	6 (197)	22, 26, 27, 29, 31, 32
	No intervention	9 (195)	21, 23, 25, 26, 28, 31, 33-35
Comparison	Nonsimulation training	4 (47)	22, 29, 36, 37
	Alternate simulation training	4 (147)	24, 27, 30, 32
	Medical students	1 (20)	24
Participants <sup>b</sup>	Physicians postgraduate training	13 (269)	21-23, 26-28, 30, 31, 33-37
	Physicians in practice	2 (63)	27, 32
	Other/ambiguous/mixed	3 (37)	25, 27, 29
Task <sup>b</sup>	Flexible bronchoscopy	16 (369)	21, 22, 24-37
	Rigid bronchoscopy/foreign body	2 (37)	23, 35
	Endobronchial ultrasound	2 (21)	36, 37
	Transbronchial needle aspiration	2 (56)	32, 37
Simulation modalities <sup>b</sup>	Virtual reality	12 (239)	21, 22, 24-26, 29, 30, 32, 33, 35-37
	Part-task model	6 (195)	27, 28, 31, 32, 34, 35
	Animal tissue	2 (56)	23, 30
	Mannequin	2 (53)	30, 35
Outcomes <sup>b</sup>	Satisfaction	3 (84)	24, 29, 32
	Knowledge	2 (86)	33, 34
	Skill: time	7 (147)	21, 23, 25, 27, 28, 31, 35, 36
	Skill: process	10 (205)	21, 23-25, 27, 28, 30, 34-36
	Behavior: time	3 (24)	22, 26, 37
	Behavior: process	4 (68)	22, 26, 33, 37
	Patient effects	1 (8)	37
Quality	Newcastle-Ottawa $\geq$ 4 points	4 (99)	22, 26, 27, 31
	MERSQI $\geq$ 12 points	7 (196)	22, 26, 27, 31, 33, 34, 37

See Table 1 legend for expansion of abbreviation.

<sup>a</sup>Numbers reflect the number enrolled, except for Outcomes, which reflects the number of participants who provided observations for analysis.

<sup>b</sup>The number of studies and trainees in some subgroups may sum to more than the number for all studies because several studies included > 1 trainee group or simulation modality, fit within > 1 task, or reported > 1 outcome.

### Synthesis

**Effectiveness in Comparison With No Training:** Nine studies compared simulation-based training with no intervention (either single-group comparison with baseline, or in comparison with a control group). For the eight studies reporting process outcomes (technique or completeness; six tested with a simulator, two tested in real patients), the pooled effect size was 1.21 (95% CI, 0.82-1.60;  $P < .0001$ ) (Fig 2A). This is considered a large effect size favoring simulation training. The outcomes were moderately consistent between studies, with  $I^2 = 36\%$  and individual effect sizes ranging from 0.68 to 3.85. The effect size for the one randomized trial was 1.54.<sup>26</sup> When restricting the analysis to the two studies with outcomes in real patients,<sup>26,33</sup> the effect size was slightly smaller (0.86 [95% CI, 0.18-1.55;  $P = .01$ ]) but still large in magnitude. The funnel plot was symmetric, and sensitivity analysis excluding studies with imprecise effect size revealed a similar pooled result.

When looking at the seven studies reporting time outcomes (six tested with a simulator, one tested in real patients), the pooled effect size was 0.62 (95% CI, 0.12-1.13;  $P = .02$ ), with a large  $I^2 = 55\%$

(Fig 2B). The one study of time behaviors with real patients found that training actually increased procedure time,<sup>26</sup> but was accompanied by a significant improvement in process (technique). The remaining six studies (all tested in a simulator) all showed favorable effects on time. The pooled effect size for the two randomized trials was 0.56. Once again, the funnel plot was symmetric, and sensitivity analyses excluding studies with imprecise effect size revealed a similar pooled result.

**Effectiveness in Comparison With Nonsimulation Training:** Four articles reported comparisons of simulation training with nonsimulation instruction. Three of these reported comparisons with clinical training.<sup>22,36,37</sup> Of these, two articles<sup>36,37</sup> appeared to report different outcomes from the same study, so in any given meta-analysis we included only one outcome from this pair of reports. One study compared intense virtual-reality simulation (five different cases, approximately four repetitions on each case, spread over three to four training sessions) with clinical training (10-15 bronchoscopies in a community hospital) using behavior outcomes.<sup>22</sup> The other also

**Table 3—Summary of Quality of Included Studies**

Scale Item	Subscale (Points if Present)	No. (%) Present; N = 17
<b>MERSQI<sup>a</sup></b>		
Study design (maximum 3)	One-group pre-post (1.5)	6 (35)
	Observational two-group (2)	5 (30)
	Randomized two-group (3)	6 (35)
Sampling: No. institutions (maximum 1.5)	1 (0.5)	13 (76)
	2 (1)	0
	> 2 (1.5)	4 (24)
Sampling: Follow-up (maximum 1.5)	< 50% or not reported (0.5)	5 (30)
	50%-74% (1)	0
	≥ 75% (1.5)	12 (70)
Type of data: Outcome assessment (maximum 3)	Subjective (1)	3 (18)
	Objective (3)	14 (82)
Validity evidence (maximum 3)	Content (1)	3 (18)
	Internal structure (1)	1 (6)
	Relations to other variables (1)	3 (18)
Data analysis: appropriate (maximum 1)	Appropriate (1)	13 (76)
Data analysis: sophistication (maximum 2)	Descriptive (1)	2 (12)
	Beyond descriptive analysis (2)	15 (88)
Highest outcome type (maximum 3)	Reaction (1)	2 (12)
	Knowledge, skills (1.5)	11 (65)
	Behaviors (2)	3 (18)
	Patient/health-care outcomes (3)	1 (6)
<b>Newcastle-Ottawa Scale (Modified)<sup>b</sup></b>		
Representativeness of sample	Present (1)	1 (6)
Comparison group from same community	Present (1)	10 (59)
Comparability of comparison cohort, criterion A <sup>c</sup>	Present (1)	6 (35)
Comparability of comparison cohort, criterion B <sup>c</sup>	Present (1)	1 (6)
Blinded outcome assessment	Present (1)	10 (59)
Follow-up high	Present (1)	12 (70)

See Table 1 legend for expansion of abbreviation.

<sup>a</sup>Mean (SD) MERSQI score was 11.4 (2.0); median (range) was 11 (9-16).

<sup>b</sup>Mean (SD) Newcastle-Ottawa Scale score was 2.4 (1.5); median (range) was 2 (0-5).

<sup>c</sup>Comparability of cohorts criterion A was present if the study (1) was randomized, or (2) controlled for a baseline learning outcome; criterion B was present if (1) a randomized study concealed allocation, or (2) an observational study controlled for another baseline trainee characteristic.

compared virtual-reality simulation (15 cases) with clinical training (15-25 real patients), using outcomes of skills<sup>36</sup> and actions with real patients (time, technique, and procedural success).<sup>37</sup> We summarize these results in Figure 3. Meta-analyses for time and process behaviors with real patients<sup>22,37</sup> showed moderate effects favoring simulation for time (0.61 [95% CI, -1.47 to 2.69]) and small effects for process (0.33 [95% CI, -1.46 to 2.11]) but confidence intervals were wide and did not exclude the possibility of no effect ( $P \geq .57$ ). One study each reported outcomes of time skill,<sup>36</sup> process skill,<sup>36</sup> and effects on patients<sup>37</sup>; all effect sizes were large (all > 0.80) and favored simulation, but only the difference for time was statistically significant. Notably, training time and structure varied between the simulation and clinical training, and these differences may account in part for the observed effects.

The fourth article reported the comparison of hands-on virtual-reality simulation training vs watching videos of bronchoscopy.<sup>29</sup> The simulation group had much higher satisfaction; no other outcomes were assessed.

*Features of Effective Instructional Design:* For evidence-based education to occur, teachers need guidance on key decisions in course development such as feedback (how and how much to deliver), repetitions (how many and in what sequence), distribution of practice (1 day or spread over several days), and authenticity (need for fidelity or accurate clinical context).<sup>20</sup> We sought evidence for these and other instructional design features in an inductive analysis of all studies making comparison with other active instruction (ie, comparative-effectiveness studies). The four studies making direct comparisons between different simulation-based approaches offered the most straightforward evidence in this regard,<sup>24,27,30,32</sup> but we also considered comparisons with nonsimulation training. We found three dominant themes: simulation modality, duration and structure of training, and clinical context.

Three studies compared different simulation modalities. One study compared a plastic model with a choose-the-hole box (see description in the “Study Characteristics” section).<sup>27</sup> After 2 weeks of self-guided practice with their respective model, participants

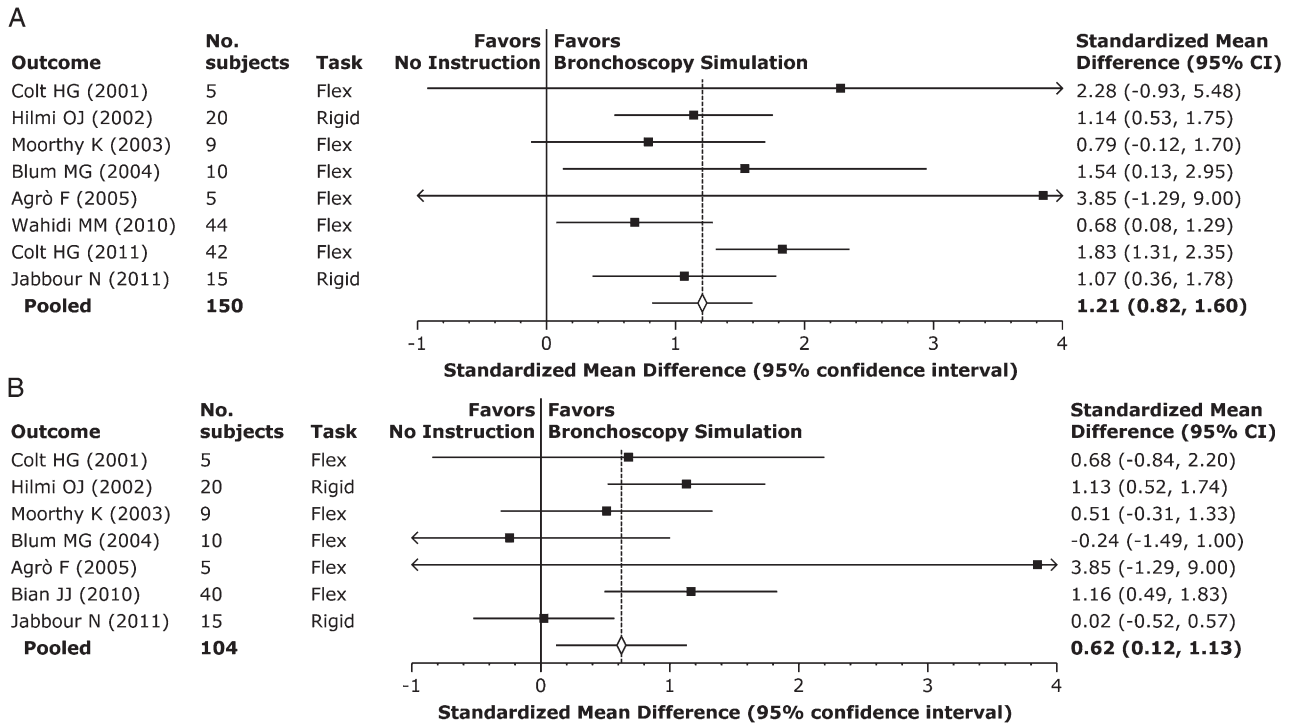


FIGURE 2. Meta-analysis: simulation-based training compared with no intervention. A, Outcome: process skills and behaviors. B, Outcome: time skills and behaviors.

performed bronchoscopy on another course participant. Speed (time) and technique (process skills) were superior for those trained with the plastic model. Those with the plastic model also reported spending more time practicing and rated the model as more interesting and useful. A crossover study found higher satisfaction and perceived usefulness for a rubber-plastic model in comparison with a virtual-reality simulator when learning transbronchial needle aspiration.<sup>32</sup> Finally, one study evaluated a multimodal course including a lecture, animal laboratory, high-fidelity mannequin, human standardized patient, and virtual-reality simulator.<sup>30</sup> When asked to compare these in terms of impact on skills, participants rated the animal laboratory as most useful for nearly all metrics, the mannequin in second place, and the virtual-reality simulator as least useful. While it is difficult to draw clear conclusions from this disparate

body of evidence, it seems that virtual-reality simulation may be less useful than other modalities, and that more interesting models may stimulate learners to practice more.

Another three studies compared variations in the structure and duration of training. In the comparisons of simulation training with clinical training, simulation was superior to clinical when the simulation training was highly structured (20 different cases, multiple repetitions per case, and spread over several practice sessions), and clinical training was informal and cases were fewer in number (10-15 unselected cases).<sup>22</sup> By contrast, behavior outcomes slightly favored clinical training when the clinical training was more structured and the number of practice cases was intentionally similar to those in the simulation group.<sup>36,37</sup> A third study (discussed in the preceding paragraph) found superior skills after longer practice with a plastic model

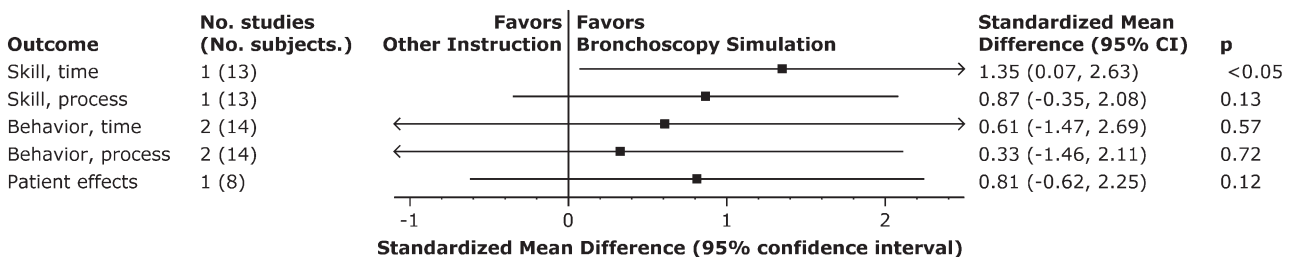


FIGURE 3. Synthesis: simulation-based training compared with clinical training. Results for two studies are pooled using meta-analysis; results for single studies are simply the effect size.

compared with shorter practice with a wooden choose-the-hole box.<sup>27</sup> Each of these comparisons is, unfortunately, confounded with concurrent differences in the modality (simulation vs clinical, or plastic model vs wooden box), so we cannot with certainty attribute the observed differences to variation in the structure, duration/repetitions, or modality.

One study compared virtual-reality simulation with or without an authentic clinical context (a written case scenario and use of surgical clothing). This study found a small, nonsignificant improvement in skills favoring the authentic context.<sup>24</sup>

## DISCUSSION

This systematic review identified 17 studies evaluating simulation-based bronchoscopy training in comparison with no intervention or alternate instruction. In comparison with no intervention, simulation-based training was consistently associated with better learning outcomes. Pooled effect sizes were large for process outcomes and moderate for time outcomes, including several studies reporting behaviors with real patients. In comparison with nonsimulation instruction, simulation-based training was again associated with higher outcomes, although these differences were generally not statistically significant and could be caused by factors other than the simulation itself.

We also conducted a narrative evidence synthesis in search of guidance for designing simulation-based bronchoscopy training. Comparisons of different simulation modalities demonstrated that a plastic part-task model representing the human bronchial tree was more effective and engendered greater satisfaction than a nonanatomic wooden model, and that a rubber-plastic model and animal model were favored over a virtual-reality simulator. Three comparisons of different training approaches suggested that longer or more structured training is more effective. One study suggested that an authentic clinical context could lead to improved skills. However, all of these findings must be considered preliminary due to confounding (unclear source of effect) or failure to reach statistical significance.

### *Limitations and Strengths*

As in any review, our results are limited by the quantity and quality of the original studies. Although one-third of the studies were randomized trials, another third were single-group pre-post studies. Follow-up was low in 30%, outcomes were blinded in fewer than 60%, and validity evidence was infrequently reported. However, most outcomes were objectively determined, four studies were multi-institutional, and three studies

reported patient-related outcomes. Overall, the study quality was similar to that in previous reviews.<sup>12,38</sup>

Between-study inconsistency was moderate for process outcomes and high for time outcomes. We suspect this was due in large part to variation in simulation modalities, outcome measures, and to a lesser extent trainees. While this diversity is a weakness in terms of between-study heterogeneity, it is also a strength in terms of comprehensiveness and breadth of scope. Moreover, all but one outcome favored simulation-based training in comparison with no intervention, and five of seven outcomes favored simulation training in comparison with nonsimulation instruction, indicating that studies generally varied in the magnitude but not the direction of benefit.

We found few studies making comparison with active interventions. This limits the strength of the inferences regarding features of effective instructional design.

Our review has several strengths, including a comprehensive literature search, rigorous and reproducible coding, and focused analyses. We included two articles published in non-English languages. Our results were robust to sensitivity analyses excluding low-quality studies, and we did not detect obvious publication bias.

### *Comparison With Previous Reviews*

Two previous reviews of simulation-based bronchoscopy training offered a detailed description of the modalities and instructional designs.<sup>10,11</sup> To these reviews we add 10 new studies identified in a systematic search, a rigorous evaluation of study quality, and quantitative synthesis of results using meta-analysis. Our results agree in general with previous reviews of simulation-based medical education<sup>6,12,39,40</sup> in that studies making comparison with no intervention typically find large effects.

### *Implications*

Our findings have important implications for current practice and future research. Simulation-based bronchoscopy training is effective for a variety of tasks including inspection, foreign-body removal, endobronchial ultrasound, transbronchial needle aspiration, and rigid bronchoscopy, as well as learning transfers to patient care. Educators should consider using this approach when resources permit. Simulation-based training is complemented by simulation-based assessment, done using tools such as the Bronchoscopy Skills and Tasks Assessment Tool.<sup>41</sup> Together, training and assessment lend themselves to competency-based education<sup>42</sup> and mastery learning.<sup>43</sup> None of the studies in our sample used a mastery learning model, in which trainees must reach a specified level of proficiency

before advancing to the next phase of instruction. Given the benefits of this model in training other procedural skills,<sup>44</sup> educators might consider mastery learning when training for bronchoscopy.

Few studies focused on advanced skills such as rigid bronchoscopy, transbronchial biopsy, or pediatric populations. While these few studies demonstrated results concordant with the larger sample, such topics may merit additional attention, particularly since these are the tasks that fellows see less often during their clinical training.<sup>9</sup>

Finally, while simulation-based training is clearly effective, the optimal design of such instruction is much less clear. The present body of evidence provides only weak support for important instructional-design decisions. It appears that longer or more structured training adds value, and that animal models and mannequins may be superior to (and possibly less expensive than) virtual-reality simulators. However, virtual reality offers other advantages such as intentional sequencing of case topics and difficulty, automated scoring and feedback, and unrestricted availability. We found little evidence to guide researchers regarding the provision of feedback, the number and sequence of training tasks, or the choice of modalities. Since differences in instructional design and assessment have potentially substantial implications for learning efficiency and optimal use of faculty time, researchers should seek to clarify such decisions<sup>45</sup> in future research.

#### ACKNOWLEDGMENTS

**Author contributions:** Dr Cook had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Dr Kennedy:** contributed by acquiring data, drafting the manuscript, and critically revising the manuscript for important intellectual content.

**Dr Maldonado:** contributed by acquiring data, drafting the manuscript, and critically revising the manuscript for important intellectual content.

**Dr Cook:** contributed to the concept and design of the study; acquired, analyzed, and interpreted data; drafted the manuscript; performed statistical analysis; obtained funding; provided administrative, technical, or material support; supervised the study; and critically revised the manuscript for important intellectual content.

**Financial/nonfinancial disclosures:** The authors have reported to CHEST that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

**Role of sponsors:** 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.

**Other contributions:** We thank Ryan Brydges, PhD; Patricia J. Erwin, MLS; Stanley J. Hamstra, PhD; Rose Hatala, MD; Jason H. Szostek, MD; Amy T. Wang, MD; and Benjamin Zendejas, MD, for their assistance in the literature search and initial data acquisition, and Essam Mekhaieel, MD, for assistance in reviewing reference lists.

#### REFERENCES

1. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med.* 2003;78(8):783-788.
2. Dong Y, Suri HS, Cook DA, et al. Simulation-based objective assessment discerns clinical proficiency in central line placement: a construct validation. *Chest.* 2010;137(5):1050-1056.
3. Issenberg SB, McGaghie WC, Hart IR, et al. Simulation technology for health care professional skills training and assessment. *JAMA.* 1999;282(9):861-866.
4. Ovassapian A, Yelich SJ, Dykes MH, Golman ME. Learning fibreoptic intubation: use of simulators v. traditional teaching. *Br J Anaesth.* 1988;61(2):217-220.
5. Ti LK, Chen F-G, Tan G-M, et al. Experiential learning improves the learning and retention of endotracheal intubation. *Med Educ.* 2009;43(7):654-660.
6. Ma IW, Brindle ME, Ronksley PE, Lorenzetti DL, Sauve RS, Ghali WA. Use of simulation-based education to improve outcomes of central venous catheterization: a systematic review and meta-analysis. *Acad Med.* 2011;86(9):1137-1147.
7. Wayne DB, Butter J, Siddall VJ, et al. Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. *J Gen Intern Med.* 2006;21(3):251-256.
8. Haponik EF, Russell GB, Beamis JF Jr, et al. Bronchoscopy training: current fellows' experiences and some concerns for the future. *Chest.* 2000;118(3):625-630.
9. Pastis NJ, Nietert PJ, Silvestri GA; American College of Chest Physicians Interventional Chest/Diagnostic Procedures Network Steering Committee. Variation in training for interventional pulmonary procedures among US pulmonary/critical care fellowships: a survey of fellowship directors. *Chest.* 2005;127(5):1614-1621.
10. Davoudi M, Colt HG. Bronchoscopy simulation: a brief review. *Adv Health Sci Educ Theory Pract.* 2009;14(2):287-296.
11. Stather DR, Lamb CR, Tremblay A. Simulation in flexible bronchoscopy and endobronchial ultrasound: a review. *J Bronchology Interv Pulmonol.* 2011;18(3):247-256.
12. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA.* 2011;306(9):978-988.
13. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med.* 2009;151(4):264-269. W64.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
15. 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(9):1002-1009.
16. 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(10):1181-1196.
17. Kirkpatrick D. Great ideas revisited. Techniques for evaluating training programs. Revisiting Kirkpatrick's four-level model. *Train Dev.* 1996;50(1):54-59.
18. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557-560.
19. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
20. 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(1):10-28.

21. Colt HG, Crawford SW, Galbraith O III. Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest*. 2001;120(4):1333-1339.
22. Ost D, DeRosiers A, Britt EJ, Fein AM, Lesser ML, Mehta AC. Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med*. 2001;164(12):2248-2255.
23. Hilmi OJ, White PS, McGurty DW, Oluwole M. Bronchoscopy training: is simulated surgery effective? *Clin Otolaryngol Allied Sci*. 2002;27(4):267-269.
24. Ameer S, Carlander K, Grundström K, et al. Learning bronchoscopy in simulator improved dexterity rather than judgement [in Swedish]. *Lakartidningen*. 2003;100(35):2694-2699.
25. Moorthy K, Smith S, Brown T, Bann S, Darzi A. Evaluation of virtual reality bronchoscopy as a learning and assessment tool. *Respiration*. 2003;70(2):195-199.
26. Blum MG, Powers TW, Sundaresan S. Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *Ann Thorac Surg*. 2004;78(1):287-291.
27. Martin KM, Larsen PD, Segal R, Marsland CP. Effective nonanatomical endoscopy training produces clinical airway endoscopy proficiency. *Anesth Analg*. 2004;99(3):938-944.
28. Agrò F, Sena F, Lobo E, Scarlata S, Dardes N, Barzoi G. The Dexter Endoscopic Dexterity Trainer improves fiberoptic bronchoscopy skills: preliminary observations. *Can J Anaesth*. 2005;52(2):215-216.
29. Chen J-S, Hsu H-H, Lai I-R, et al; National Taiwan University Endoscopic Simulation Collaborative Study Group (NTUSEC). Validation of a computer-based bronchoscopy simulator developed in Taiwan. *J Formos Med Assoc*. 2006;105(7):569-576.
30. Deutsch ES, Christenson T, Curry J, Hossain J, Zur K, Jacobs I. Multimodality education for airway endoscopy skill development. *Ann Otol Rhinol Laryngol*. 2009;118(2):81-86.
31. Bian JJ, Wang JF, Wan XJ, Zhu KM, Deng XM. Preparation of a fiberoptic bronchoscopy training box and evaluation of its efficacy [in Chinese]. *Academic Journal of Second Military Medical University*. 2010;31:80-83.
32. Davoudi M, Wahidi MM, Zamanian Rohani N, Colt HG. Comparative effectiveness of low- and high-fidelity bronchoscopy simulation for training in conventional transbronchial needle aspiration and user preferences. *Respiration*. 2010;80(4):327-334.
33. Wahidi MM, Silvestri GA, Coakley RD, et al. A prospective multicenter study of competency metrics and educational interventions in the learning of bronchoscopy among new pulmonary fellows. *Chest*. 2010;137(5):1040-1049.
34. Colt HG, Davoudi M, Murgu S, Zamanian Rohani N. Measuring learning gain during a one-day introductory bronchoscopy course. *Surg Endosc*. 2011;25(1):207-216.
35. Jabbour N, Reihsen T, Sweet RM, Sidman JD. Psychomotor skills training in pediatric airway endoscopy simulation. *Otolaryngol Head Neck Surg*. 2011;145(1):43-50.
36. Stather DR, Maceachern P, Rimmer K, Hergott CA, Tremblay A. Assessment and learning curve evaluation of endobronchial ultrasound skills following simulation and clinical training. *Respirology*. 2011;16(4):698-704.
37. Stather DR, Maceachern P, Chee A, Dumoulin E, Tremblay A. Evaluation of clinical endobronchial ultrasound skills following clinical versus simulation training. *Respirology*. 2012;17(2):291-299.
38. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. *Acad Med*. 2010;85(5):909-922.
39. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med*. 2011;86(6):706-711.
40. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med*. 2010;85(10):1589-1602.
41. Davoudi M, Osann K, Colt HG. Validation of two instruments to assess technical bronchoscopic skill using virtual reality simulation. *Respiration*. 2008;76(1):92-101.
42. Weinberger SE, Pereira AG, Iobst WF, Mechaber AJ, Bronze MS; Alliance for Academic Internal Medicine Education Redesign Task Force II. Competency-based education and training in internal medicine. *Ann Intern Med*. 2010;153(11):751-756.
43. McGaghie WC, Siddall VJ, Mazmanian PE, Myers J; American College of Chest Physicians Health and Science Policy Committee. Lessons for continuing medical education from simulation research in undergraduate and graduate medical education: effectiveness of continuing medical education: American College of Chest Physicians evidence-based educational guidelines. *Chest*. 2009;135(suppl 3):62S-68S.
44. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med*. In press. doi:10.1097/ACM.0b013e31829a365d.
45. Cook DA, Bordage G, Schmidt HG. Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Med Educ*. 2008;42(2):128-133.