Retention of Critical Procedural Skills After Simulation Training: A Systematic Review

Camille Legoux, MD, CM[1], Richard Gerein, MD[2,3,4], Kathy Boutis, MD, MSc[5], Nicholas Barrowman, PhD[2,3], and Amy Plint, MD, MSc[2,3,4]

ABSTRACT

Objective: While short-term gains in performance of critical emergency procedures are demonstrated after simulation, long-term retention is relatively uncertain. Our objective was to determine whether simulation of critical emergency procedures promotes long-term retention of skills in nonsurgical physicians.

Methods: We searched multiple electronic databases using a peer-reviewed strategy. Eligible studies 1) were observational cohorts, quasi-experimental or randomized controlled trials; 2) assessed intubation, cricothyrotomy, pericardiocentesis, tube thoracostomy, or central line placement performance by nonsurgical physicians; 3) utilized any form of simulation; and 4) assessed skill performance immediately after and at ≥ 3 months after simulation. The primary outcome was skill performance at or above a preset performance benchmark at ≥ 3 months after simulation. Secondary outcomes included procedural skill performance at 3, 6, and ≥ 12 months after simulation.

Results: We identified 1,712 citations, with 10 being eligible for inclusion. Methodologic quality was moderate with undefined primary outcomes; inadequate sample sizes; and use of nonstandardized, unvalidated tools. Three studies assessed performance to a specific performance benchmark. Two demonstrated maintenance of the minimum performance benchmark while two demonstrated significant skill decay. A significant decline in the mean performance scores from immediately after simulation to 3, 6, and ≥ 12 months after simulation was observed in four of four, three of four, and two of five studies, respectively. Scores remained significantly above baseline at 3, 6, and ≥ 12 months after simulation in three of four, three of four, and four of four studies, respectively.

Conclusion: There were a limited number of studies examining the retention of critical skills after simulation training. While there was some evidence of skill retention after simulation, overall most studies demonstrated skill decline over time.

A physician's ability to perform critical emergency procedures during resuscitation is a high-stakes and potentially lifesaving skill. However, since these procedures are rarely performed in the context of standard clinical practice, physicians' exposure can be limited. This is likely even more true in pediatric settings where critical procedures are less frequent that in general emergency department (ED) settings.[1,2] For

From [1]Harvard Medical School, Boston, MA; the [2]Children's Hospital of Eastern Ontario (CHEO); and the [3]Department of Pediatrics; and the [4]Department of Emergency Medicine, University of Ottawa, Ottawa, Ontario, Canada; and [5]The Hospital for Sick Children and Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada.
Received July 11, 2020; revision received September 2, 2020; accepted September 11, 2020.
CL received peer-reviewed grant support for this project from the CHEO Research Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. AP is supported in part by a University of Ottawa Research Chair in Pediatric Emergency Medicine. The other authors have no relevant financial information or potential conflicts to disclose.
Author contributions: CL—study concept and design, acquisition of the data, analysis and interpretation of the data, drafting of the manuscript, acquisition of funding; RG—acquisition of the data, critical revision of the manuscript for important intellectual content; KB—study concept and design, critical revision of the manuscript for important intellectual content; NB—analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, statistical expertise; and AP—study concept and design, analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, acquisition of funding.
Supervising Editor: Anne Messman, MD.
Address for correspondence and reprints: Camille Legoux, MD, CM; e-mail: camille.legoux@mail.mcgill.ca.
AEM EDUCATION AND TRAINING 2021;5:1–12.

doi: 10.1002/aet2.10536

example, endotracheal intubation is performed by physicians in 0.55% of adult ED visits,[1] a frequency that is 4.4 times higher than among pediatric ED visits.[3,4] As a result, it is not surprising that most pediatric emergency physicians report not having performed a single intubation over a 12-month period,[3] and while endotracheal intubation is the most commonly performed critical emergency procedure in both adults and children, other procedures are even much less commonly performed. For example, pericardiocentesis is performed in only 0.007% of emergency visits by adults,[1] a frequency 8.3-fold higher than among pediatric ED visits.[3] Nevertheless, given the low frequency at which emergency situations requiring critical procedure skills are encountered, it is unlikely that physicians can maintain the necessary skills and competence through exposure alone in both general and pediatric ED settings.

By providing additional exposure and training opportunities, simulation can serve a crucial function in preparing clinicians to encounter critically ill patients and rare events.[5] However, simulation training is not without limitations. While gains in both procedural knowledge and kinesthetic skills have been demonstrated immediately following a simulation exercise, there is limited evidence on the long-term retention of these rare but crucial skills. A systematic review on advanced life support skills for health care providers showed that the most significant decline in skill occurred between 6 and 12 months after course completion.[6] The retention of other procedural skills has been less extensively studied, but supports that simpler procedural skills tend to be more resistant to decay than complex ones.[7,8] However, an accurate representation retention of procedural skills after simulation exercises may be limited due to methodologic issues in the existing literature of individual studies such as small sample size of participants, heterogeneity in teaching methods, and assessments without validity evidence.[6,9] Nevertheless, an accurate longitudinal representation of the relationship between skill acquisition and decay over time is important to be able to strategically schedule skill practice and assessment which ensures the durability of competence over time.[10] A systematic review with meta-analysis examining the learning retention of critical procedural skills after simulation would allow a formal assessment of the quality of evidence and a quantitative comparison of outcomes and may emphasize the key knowledge gaps that need to be answered to better define education solutions.

The main objective of our study was to determine whether critical emergency skill performance can be maintained at or above a preset performance benchmark at different time points after the initial simulation learning intervention. The secondary objective was to determine the change in performance at different time points. We hypothesized that there would be a substantial decay in skill performance as early as 3 months after simulation.

METHODS

Study Design

We performed a systematic review with meta-analysis to answer the question: Can critical emergency skill performance be maintained at or above a preset performance benchmark at least 3 months after simulation training?

Study Eligibility

Eligibility criteria included study design, study population, critical emergency procedure skill performance in the simulation setting, and measurement of procedural skill performance in the simulation setting. All publications needed to be full length and peer reviewed. For study design, we considered observational cohort, quasi-experimental, or randomized controlled trials (RCTs). Studies that examined the retention of skill performance for multistep critical emergency procedures by nonsurgical physicians (residents, fellows, and attending physicians) were considered for inclusion. Physicians practicing either pediatric or adult medicine were included. The procedure studied was required to be one of the following five multistep critical emergency procedures: intubation, cricothyrotomy, pericardiocentesis, tube thoracostomy, or central line placement.

The assessment of skill performance in the study needed to 1) be performed objectively, using tools such as a performance checklist; and 2) consider the checklist score obtained as representative of the participants' procedural skill performance. For studies comparing participants' procedural skill performance to a reference standard, the performance benchmark needed to represent a minimum preset level or threshold to be met by participants. Consequently, obtaining a score inferior to this performance benchmark would

categorize the performance as unsatisfactory (failing to meet the minimum standard).

Eligible studies could either compare the performance of a single group at two distinct time periods or compare the performance of an intervention group (who received simulation training) to that of a control group (who did not receive simulation training). Performance skills had to be measured immediately after simulation training and at least 3 months after simulation training. Measurement of performance prior to simulation training was not required to meet inclusion criteria. The 3-month interval was used as a surrogate for long-term retention. Studies were not excluded based on the realism of the simulation method.

We excluded studies that examined critical emergency procedure skill performance by medical students, physicians from surgical specialties, and allied health care personnel. We also excluded studies that exclusively examined communication and team dynamics or known resuscitation algorithms (e.g., neonatal resuscitation program, pediatric advanced life support, advanced cardiac life support).

Literature Search

The main search strategy (see Data Supplement S1, Table S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10536/full) was developed by a medical librarian in consultation with the research team and content experts. The search strategy was peer-reviewed by a second medical librarian. We conducted a systematic search of the literature from inception to February 1, 2019. We searched the following bibliographic databases: MEDLINE, including Epub ahead of print, in-process, and other nonindexed citations (1946 to February 1, 2019), Embase (1947 to February 2019), and the CENTRAL Trials Registry of the Cochrane Collaboration (January 2019 issue) using the Ovid interface. Search terms included words and phrases associated with critical emergency procedures in concert with the terms "simulation," "skill performance," and "retention." There was no restriction on publication language.

Study Selection

After initial online duplicate removal, records retrieved by the electronic search were downloaded and imported into a Reference Manager database, where any remaining duplicate references were removed. Records were uploaded to InsightScope and appraised against the inclusion criteria using a three-step approach. First, two reviewers independently reviewed the title and abstract of each potentially relevant article identified from the search and classified them as "included" or "not included." If an abstract was deemed included by one reviewer, the full article was retrieved. The same two reviewers then independently evaluated the full-length articles for inclusion. Disagreements were resolved by consensus between the two reviewers.

Data Extraction

Data were abstracted using a standardized form by one author, with the second author verifying extraction accuracy. Abstracted variables included study design, critical emergency skill studied, participants (level of training, discipline and number), realism of simulation experience, description of the simulation intervention, method of performance evaluation, performance evaluation (baseline performance, immediately after intervention, and at $\geq$ 3 months after simulation), definition of the minimum preset performance benchmark of procedural skill performance, and funding source.

Assessment of Methodologic Quality

One author assessed the methodologic quality of each study, with accuracy verified by the second author. We rated the methodologic quality of included studies using the Medical Education Research Study Quality Instrument (MERSQI).[11,12]

Study Outcomes

The primary outcome for the review was defined as demonstration of skill performance at or above a minimum preset performance benchmark at $\geq$ 3 months after simulation. Secondary outcomes included procedural skill performance at 3, 6, and $\geq$ 12 months after simulation.

Data Analysis

We performed both qualitative and quantitative analyses. We considered each assessment time point for each procedure as a separate study group. For each group of each included article, the mean checklist score and standard deviation (SD) were extracted for the following time points: baseline (i.e., prior to simulation intervention), immediately after simulation, and at the delayed after simulation assessment (3, 6, and/

or ≥ 12 months). Raw scores were transformed into percentages and SDs. If not reported directly, transformed scores were calculated based on reported confidence intervals and interquartile ranges. Since the individual scores of participants were unknown, we assumed a within-group score correlation of 0.5 to calculate the SD for each of the included studies.[13] The mean change from baseline score to delayed after simulation score (3, 6, and/or ≥ 12 months) and the mean change from immediately postsimulation score to delayed postsimulation score (3, 6, and/or ≥ 12 months) were calculated, as were the corresponding SDs. Because we expected substantial heterogeneity between study estimates, we decided to report the pooled variance pooled results only if the I^2 was less than 80%. In addition to the global analysis of change in skills performance at the three time points of interest, we performed group analyses for each procedure.

RESULTS

Study Selection

Our search strategy identified 2,573 potentially relevant citations (Figure 1). Screening of titles and abstracts excluded 1,631 citations, 81 full-text articles were assessed for eligibility, and 10 publications met inclusion criteria.

Study Characteristics and Methodologic Quality

The included articles included seven cohort studies,[2,14–19] two RCTs,[20,21] and one crossover study (Table 1).[22] Methodologic quality of the included studies was moderate with weakness in the number of institutions sampled; validation of evaluation instrument in relationship to other variables, such as the novice-to-expert transition; and level of Kirkpatrick outcomes studied (Table 2). Other areas of weakness included study group selection and comparability for the cohort studies and weakness in selective reporting and blinding for the RCTs.

All included articles were published between 2010 and 2016. A total of 317 participants were included in the 10 studies, with a median of 32 participants per study (range = 12 to 52 participants). Study participants were physicians from a wide variety of specialties. The majority of included participants were residents at different stages in their postgraduate training.[14–18,21,22] Only three studies assessed attending physicians.[19,20] The majority of included studies (seven of 10) assessed central venous line catheter insertion skill. No study examining either pericardiocentesis or tube thoracostomy met the inclusion criteria.

For each study the research teams designed an educational intervention including a didactic session, hands-on practice with a manikin, and a period for feedback. All educational interventions were designed in accordance with the deliberate practice model, a purposeful and systematic type of practice conducted with the specific goal of improving performance.[23] Additionally, two studies used mastery training, a form of competency-based education instructional strategy in which learners practice deliberately until they achieved a fixed predefined performance standard.[24] This strategy allows all learners to achieve a uniform performance, without being limited by practice time. Although instructional strategies were similar across all selected publications, the structure and breadth of these activities varied widely, as illustrated by the duration of the training sessions. For the eight studies that reported activity duration, the median duration was 90 minutes (range = 20 minutes to 4 hours). One study did not mention the duration of the educational intervention,[17] while no time was measured for the study using the mastery training strategy.[22]

All 10 included studies used a checklist to measure skill performance. Three used nonvalidated checklists,[5,6,13] two adapted checklists of known validity evidence,[18,20] and five used checklists of known validity evidence without adaptation.[2,14,17,19,21] The median number of elements on these checklists was 17 (range = 5 to 27). Each study used a consistent assessment tool for all testing; prior to simulation, immediately after simulation, and at delayed postsimulation measurement.

The timing of the delayed postsimulation evaluation ranged from 3 to 34 months. Four of the 10 studies had multiple groups of participants (such as multiple procedures or multiple timing of the delayed retention test).[14,15,20,22] Furthermore, of these four studies, two also performed repeated evaluations of delayed retention (such as a group assessed 6 months after simulation and then reassessed at 12 months after simulation).[14,22] We were able to analyze the results of 17 groups from our 10 included studies. Five groups of participants from four studies were evaluated at 3 months after simulation training,[7,9,12,13] five groups of participants from four studies were evaluated at 6 months after simulation training,[14,16,20,22] and seven

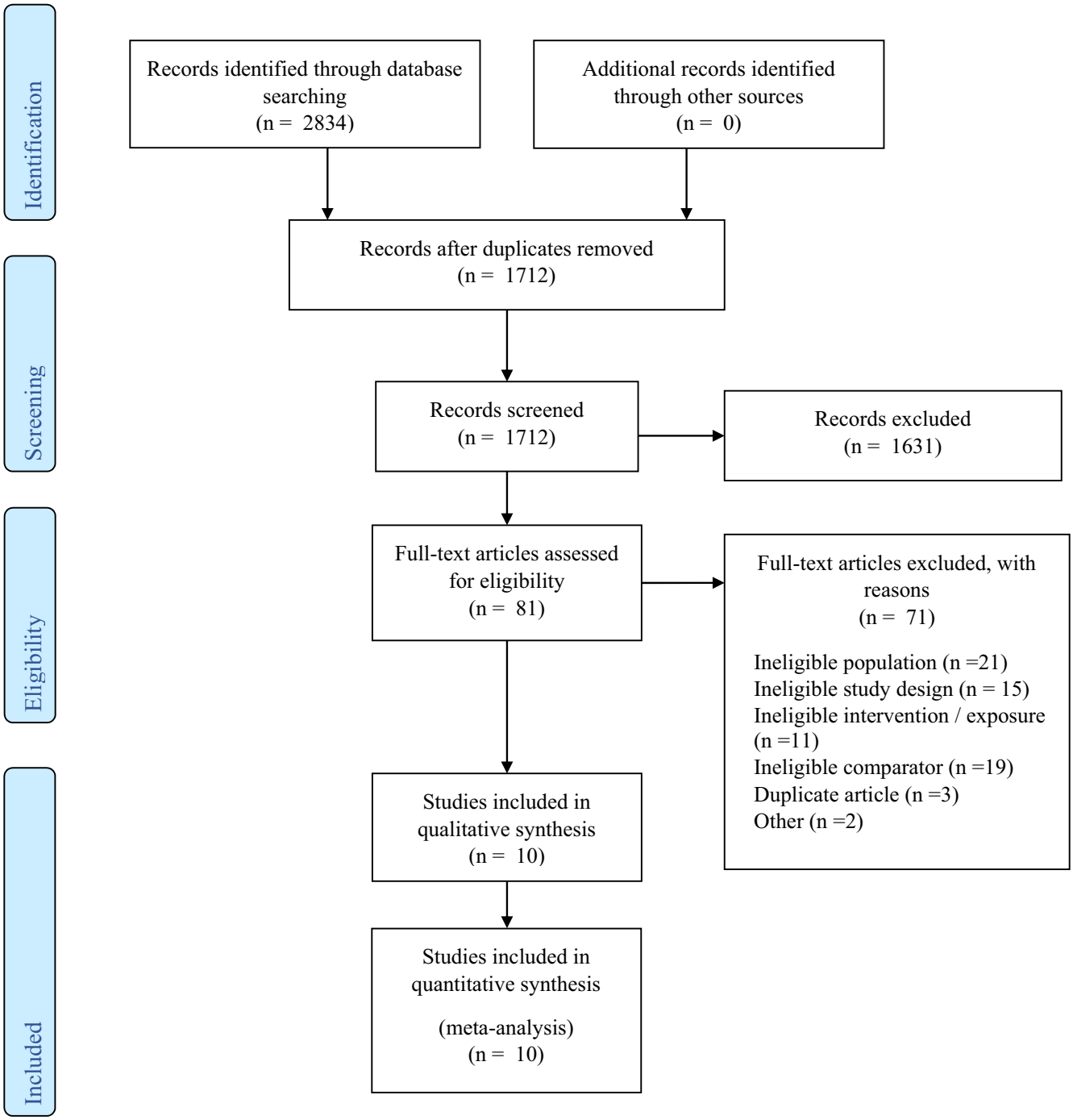

Figure 1. Flow diagram.

groups of participants from five studies were evaluated at ≥ 12 months after simulation training.[2,14,15,19,20]

Primary Outcome: Demonstration of Skill Performance at or Above a Minimum Predefined Benchmark at ≥ 3 Months After Simulation

Three of the included studies assessed skill performance using a predefined performance benchmark, representing four participants groups.[2,14,19] All were prospective cohort studies and assessed performance in central venous catheter insertion at 12 months after simulation. However, the study populations, number of delayed retention testing, and definition of the predefined performance benchmark were varied among all three studies (see Table 1). In two studies,[2,14] the predefined performance benchmark consisted of a minimum passing score of 79% on a checklist with

Table 1
Summary of Studies

Study	Study Type	Specialty and Level of Training of Participant	Number of Participants Enrolled/Study	Skill Studied	Timing of Assessment After Simulation Training	Checklist Evaluation*	Other Methods of Evaluation	Preset Performance Benchmark
Ahn 2016 United States[22]	Crossover study	Emergency medicine First-year residents	*N* = 16 (16 follow-up)	Glidescope video laryngoscopy and endotracheal intubation	3 and 6 months	Checklist score: 17 items (items scored 0 or 1) [total score 0 to 17]	Global-rating scale: 7 items (items scored 1 to 5) [total score 7 to 35] Time to completion	No
Ahya 2012 United States[2]	Cohort study	Nephrology fellows	*N* = 12 (11 follow-up)	Central venous catheter insertion (Internal jugular)	12 months	Checklist score: 27 items (items scored 0 or 1)[total score 0 to 27]†		Yes
Barsuk 2010 United States[14]	Cohort study	Internal medicine Second- and third-year residents	*N* = 36 (31 follow-up)	Central venous catheter insertion (internal jugular and subclavian)	6 and 12 months	Checklist score: 27 items (items scored 0 to 1) [total score 0 to 27]†	Global-rating scale: 8 items (items scored 1 to 5) [total score 8 to 40]	Yes
Boet 2011 Canada[20]	Single-blinded randomized study	Anesthesiology Attending physicians	*N* = 36 (34 follow-up)	Cricothyroidotomy	6 or 12 months	Checklist score: 5 items (items scored 0 to 2)[total score 0 to 10]		No
Cartier 2016 Switzerland[15]	Cohort study	Anesthesiology Second- to fifth-year residents	*N* = 37 (18 follow-up)	Central venous catheter insertion (type not specified)	More than 24 months (average 34 months)	Checklist score 1: 8 items (items scored 0 to 1)[total score 0 to 8] Checklist score 2: 20 items (scored 0 to 1) [total score 0 to 20]	Time to completion	No
Gerard 2011 United States[16]	Cohort study	Family medicine First- and second-year residents	*N* = 47 (30 follow-up)	Pediatric endotracheal intubation	6 months	Checklist score: 9 items (items scored 0 to 1) [total score 0 to 9]		No
Laack 2014 United States[17]	Cohort study	Emergency medicine Firsts- to third-year residents	*N* = 26 (23 follow-up)	Central venous catheter insertion (internal jugular)	3 months	Checklist score: 15 items (items scored 0 to 1) [total score 0 to 15]	Global-rating scale: 100-mm visual analog Time to completionNumber of needle passes	No
Smith 2010 United States[21]	RCT	Internal medicine First- and second-year residents	*N* = 52 (28 follow-up)	Central venous catheter insertion (type not specified)	3 months	Checklist score: 23 items (items scored 0 to 1) [total score 0 to 23]		No
Thomas 2013 United States[18]	Observational pilot study	Pediatrics First- to third-year residents	*N* = 27 (26 follow-up)	Central venous catheter insertion (femoral)	3 months	Checklist score: 24 items (items scored 0 to 1) [total score 0 to 24]		No
Werner 2016 United States[19]	Cohort study	Pediatric emergency medicine Attending physicians	*N* = 28 (28 follow-up)	Central venous catheter insertion (femoral)	(2 months and) 12 months	Checklist score: 17 items with 7 critical steps‡		Yes

RCT = randomized controlled trial.
*Method of evaluation that was compared immediately after simulation and at retention testing (3, 6, or ≥ 12 months).
†Competence threshold defined as reaching a minimum passing score of 79% on the checklist.
‡Competence threshold defined as performing the seven critical steps of the checklist correctly.

Table 2
Summary of Quality Assessments

MERSQI* items	Ahn[22]	Ahya[2]	Barsuk[14]	Boet[20]	Cartier[15]	Gerard[16]	Laack[17]	Smith[21]	Thomas[18]	Werner[19]
Study design	2	1	1.5	3	1.5	1.5	1.5	3	1.5	1.5
Sampling: institutions	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Sampling response rate	1.5	1.5	1.5	1.5	0.5	1	1.5	1	1.5	1.5
Type of data	3	3	3	3	3	3	3	3	3	3
Validation of instrument										
Internal structure	1	1	1	1	1	1	1	1	1	1
Content	1	1	1	1	1	1	1	1	1	1
Relationship to other variables	0	1	1	0	0	0	0	1	1	1
Data analysis										
Appropriateness	1	1	1	1	1	1	1	1	1	1
Complexity	2	2	2	2	2	2	2	2	2	2
Outcomes	1.5	2	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Total score (out of a maximum of 18)	13.5	14	14	14.5	12	12.5	13	15	14	14

*Medical Education Research Study Quality Instrument developed by Reed et al.

known validity evidence. These two studies used mastery training as a teaching strategy (so that all participants had exactly the same passing score of 100%), which precluded their inclusion in a meta-analysis with the studies that did not use this strategy.

In the study by Barsuk et al.,[14] a maintenance of skill performance at or above the predefined performance benchmark was demonstrated at 6 and 12 months after simulation. For the internal jugular and the subclavian approaches, 12.2 and 8.2% of participants, respectively, met the central venous line insertion performance benchmark prior to the simulation educational intervention, while all participants performed at or above the performance benchmark immediately after simulation. Further, 82.4 and 85.3% of the participants remained above this performance benchmark for these procedures at 6 months after simulation, while 87 and 83.9% remained above the performance benchmark 12 months after simulation (Figure 2).

In a study by Ayha et al.,[2] a loss of skill performance was demonstrated using the same checklist and minimum passing score as the study by Barsuk et al.[14] At 12 months after simulation, only 54.5% of participants remained above the performance benchmark, while all participants were above the performance benchmark immediately after simulation (Figure 2).

In a study by Werner et al.,[19] the performance benchmark was defined as the correct performance of seven critical items on a validated checklist of 17 items. This study demonstrated that 32% of the participants met the performance benchmark before simulation, and 89% achieved this benchmark immediately after simulation. At 12 months, 85% maintained this benchmark (Figure 2).

Secondary Outcomes: Procedural Skill Performance Scores at 3, 6, and ≥ 12 Months After Simulation Training

In the four studies evaluating procedural skills performance 3 months after simulation training, five groups of participants were identified. Participants were residents in their first to third years of emergency medicine, internal medicine, or pediatrics residency program. The procedure evaluated was central venous catheter insertion for three groups and endotracheal intubation using a video laryngoscope for the other two. The number of items ranged from 15 to 24 (Table 1). The mean baseline scores obtained by participants prior to simulation training were low, ranging from 54.2% to 69.9%. Scores for all groups improved immediately after simulation training and then declined significantly at 3 months after simulation training (Figure 3). At 3 months after simulation, the decrease in mean scores from immediately after simulation was statistically significant for all five groups of participants (across four studies), a decline in scores ranging from 11.0% to 34.9%.[17,18,21,22] Compared to baseline scores, however, the increase in mean scores observed at 3 months after simulation was statistically significant in four of the five groups of participants (across three studies), representing an increase of 27.3% to 36.1% in scores (Figure 3).[17,21,22]

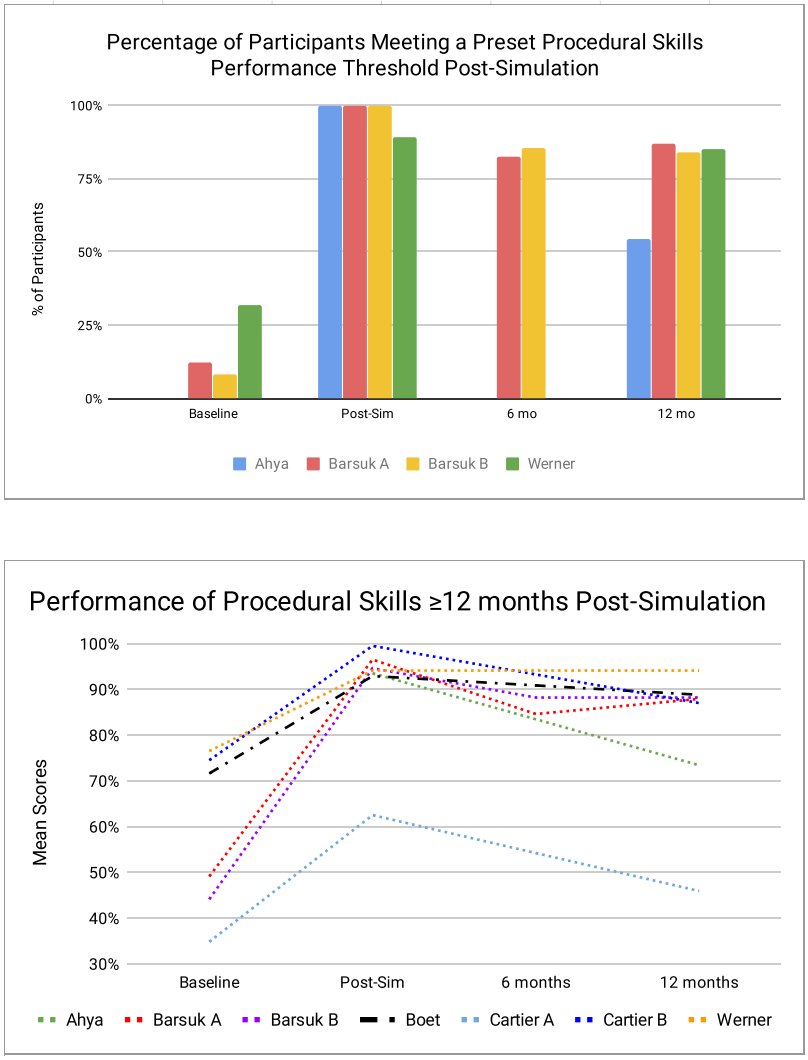

Figure 2. Participants' performance compared to a preset performance threshold at baseline, immediately post-simulation, and at 6 and ≥ 12 months post-simulation.

Similarly, for the five groups of participants (across four studies) who had a performance evaluation 6 months after simulation training, the baseline presimulation training mean scores ranged from 36.7% to 69.9% and improved immediately after simulation for all groups (Figure 3). The procedural skills evaluated in these groups consisted of endotracheal intubation with and without a video laryngoscope, central venous catheter insertion, and cricothyrotomy. The checklists used for assessment ranged from five to 27 items. The specialty and level of training of learners among these groups varied, ranging from first-year residents to attending physicians (Table 1). Compared to immediately after simulation, the decline in mean scores observed 6 months after simulation training was statistically significant for four of the five participant groups (across three studies), representing a negative percent change ranging from 6.8% to 50.2%.[14,16,20,22] Still, in four of the five groups of participants (across three studies), the mean scores increase from baseline to 6 months after simulation was also statistically significant, representing a positive percent change ranging from 22.0% to 100%.[14,20,22] (Figure 3).

Finally, for the seven groups of participants (across five studies) where performance was evaluated at ≥ 12 months after simulation training, baseline preintervention mean scores ranged from 34.8% to 76.5%. Again, these groups varied in their composition of learners' specialty and level of training. Six of the seven groups evaluated performance of central venous catheter insertion. The number of items on the checklists used for assessment ranged from five to 27 items (Table 1). The scores improved immediately after simulation training in all groups and then

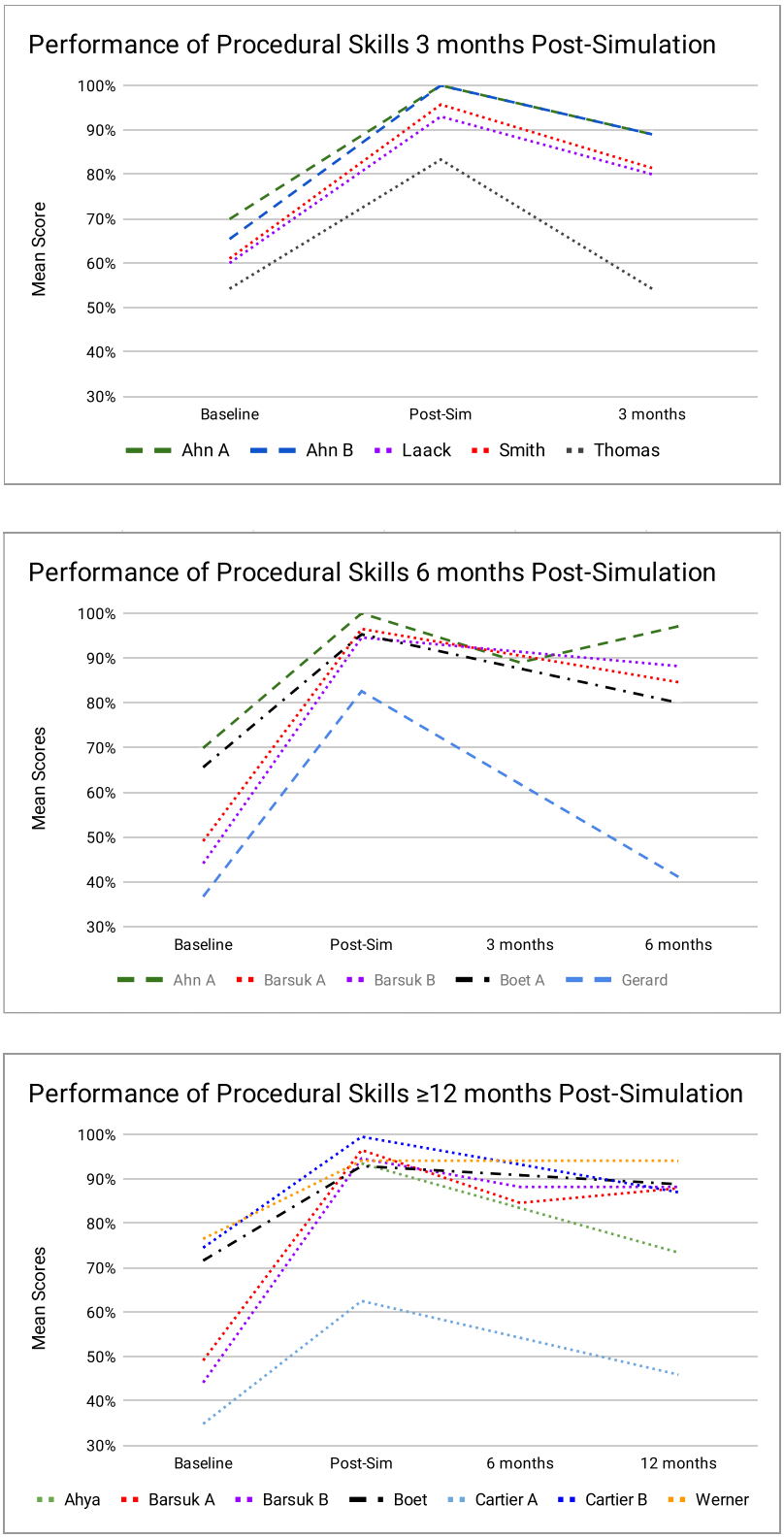

Figure 3. Participants' scores at baseline, immediately post-simulation and at 3, 6, and ≥ 12 months post-simulation. When there was high heterogeneity (I2>80%), no pooled estimate is shown.

declined at ≥ 12 months after simulation training (Figure 3). For five of the seven participant groups (across three studies), the decline in scores from immediately after simulation training to ≥ 12 months after simulation training was statistically significant, representing a negative percent change ranging from 6.8% to

26.6%.[14,15] The improvement in scores from before simulation training to ≥ 12 months after simulation training was statistically significant for all six groups of participants that had a pretraining assessment (across four studies), representing a positive percent change ranging from 16.8% to 100%[14,15,19,20] (Figure 3).

In our meta-analysis, the performance of procedural skills at the three time points of interest showed large variability (I^2 ranging from 75% to 96%), limiting the interpretation of the results. However, examining procedural skill performance from baseline to retention, the majority of studies demonstrate skill improvement (Figure 3). This can be seen in the estimates located to the right side of the forest plot axis at 3, 6, and ≥ 12 months after simulation. In addition, examining procedural skill performance from immediately after simulation to retention assessment, most studies demonstrate skills decay (Figure 3). This can be seen in the estimates located to the left side of the forest plot axis at 3, 6, and ≥ 12 months after simulation. When controlling for the timing of retention assessment and the type of procedure performed, our subgroup analysis revealed that a substantial variability remained (I^2 ranging from 53% to 94%). Results of our analysis by type of procedure can be found in Data Supplement S1 (see Figure S1 for central venous catheter insertion, Figure S2 for endotracheal intubation, and Figure S3 for cricothyrotomy).

DISCUSSION

This systematic review demonstrates that there is a statistically significant decline in mean performance scores for critical procedural skills for most of the participant groups at 3, 6, and ≥ 12 months after simulation. However, despite this decline, the mean delayed performance scores of all participant groups remained above their presimulation baseline scores. Further, of the two studies that compared the procedural skill to a predefined performance benchmark over multiple time points, the proportion of participants that remained above the performance benchmark at 12 months after simulation was about 85%,[2,14] which was a much higher proportion than reported by the study that included a single delayed postsimulation assessment.[2]

Our findings of a significant decline in critical procedural skill performance scores for most studies as early as 3 months after simulation are consistent with the existing literature on skill decay. A recent review of eight studies exploring the impact of frequency of neonatal resuscitation skills training showed that psychomotor performance skills improved when training occurred frequently, compared with annual or biennial training, supporting the fact that neonatal resuscitation program retraining should occur more frequently than the previously recommended 2-year interval.[9] In their systematic review, Yang et al.[6] found that the degree of decay of health-care providers Advanced Life Support (ALS) course skills such as management of cardiac arrest scenarios, chest compressions, and ventilation varied by study, but tended to appear between 6 months and 1 year after training. When comparing performance to a minimum performance benchmark, Yang et al. found that the proportion of health care providers meeting the performance benchmark was variable, with studies demonstrating as little as little as 14% of participants meeting the performance benchmark 12 months after training, providing evidence that ALS course skills decay significantly before the recommended interval between training and recertification. It is apparent that significant skill decay can occur early after training and that the degree of skill decay may be specific to the skill learned.

Repeated testing for assessment of retention may inform the timing of refresher education. Assessments can have an extrinsic effect on learning because of the effect of testing on motivation, learning strategies, and feedback.[25] Assessments also have an independent intrinsic effect on learning known as the testing effect, which has shown superiority as a learning strategy compared to practice alone.[26–29] The process of information retrieval from long-term memory required during a test is thought to be the main mechanism leading to strengthening of memory and testing effect. Further, repeated retrieval attempts spaced over time have been shown to be superior to a single retrieval session for the retention of knowledge and skills across a variety of domains.[30–34] For example, Park et al.[35] demonstrated that the implementation of a monthly high-yield trauma procedural training and simulation program improved resident performance as well as time to specific interventions in trauma patients in an American Level I trauma center. The finding of improved performance with retrieval practice aligns with our results whereby we demonstrated that in the two studies with multiple delayed assessments, there was superior retention relative to those with one delayed assessment.[14,22] Thus it appears that simulation, paired with retrieval practice, can promote long-term retention of some critical procedural skills, as measured by score-based assessments. However, the

form (testing only vs. complete retraining session) and optimal interval between reexposure to critical emergency skill remains to be determined. Further studies are needed to optimize the initial training, refresher education, and assessment of critical procedure to promote long-term retention of skills and minimize time, human, and financial resources.

LIMITATIONS

This review is limited by the paucity of high-quality studies in this area. There are multiple factors affecting the significance of the results we obtained. Only a small number of studies were identified for inclusion in the review, most with a small number of participants and a high loss to follow-up rate. As well, only half of the studies included in the review used checklists with known validity without further modifications.[2,14,17,19,21] The reliability and validity of the results obtained from studies utilizing other assessment instruments is unclear.[15,16,18,20,22] Further, none of the studies used evidence-based performance benchmarks, which may limit the validity of comparing scores over time. Since performance scores at the individual participant level were lacking, we chose to use a moderate correlation coefficient of 0.5 for our analysis. Although this is a likely estimate, it might not be accurate for all studies. At the retention assessment, we found a trend toward skill improvement from baseline and skill decay from immediately after simulation. However, these results obtained from our meta-analysis should be interpreted with caution given the substantial heterogeneity of the study estimates (I^2 ranging from 75% to 96%). The clinical, methodologic, and statistical heterogeneity previously described could have contributed to the degree of heterogeneity seen in the meta-analysis. Finally, although statistically and likely educationally significant, generalizability to the significance of our findings in the clinical setting is relatively unknown. Given the complexity of translating simulation-based performance to bedside performance, future studies could focus on translational simulation research to determine whether simulation interventions are effective with respect to patient outcomes.

CONCLUSION

The existing research into the simulation of critical emergency procedure is limited, with small studies having a high risk of bias and multiple flaws in design. Our results demonstrate that there is an improvement in performance scores for critical procedures from baseline to the delayed after simulation testing, but in general these scores have decreased from immediately after simulation levels. Repeated assessments may influence the degree of skill decline and would be important to consider in the planning of refresher education. Further research on critical emergency procedures simulation training and assessments should attempt to limit methodologic flaws by including larger sample size; define rigorous outcome measures including minimal performance benchmarks; consider variables such as training design, testing effect, and spacing of assessments; and aim to design translational simulation studies.

References

1. Waymack JR, Markwell S, Milbrandt JC, Clark TR. Comparison of rates of emergency department procedures and critical diagnoses in metropolitan and rural hospitals. Rural Remote Health 2015;15:3298.
2. Ahya SN, Barsuk JH, Cohen ER, Tuazon J, McGaghie WC, Wayne DB. Clinical performance and skill retention after simulation-based education for nephrology fellows. Semin Dial 2012;25:470–3.
3. Mittiga MR, Geis GL, Kerrey BT, Rinderknecht AS. The spectrum and frequency of critical procedures performed in a pediatric emergency department: implications of a provider-level view. Ann Emerg Med 2013;61:263–70.
4. Schoenfeld PS, Baker MD. Management of cardiopulmonary and trauma resuscitation in the pediatric emergency department. Pediatrics 1993;91:726–29.
5. Bradley P. The history of simulation in medical education and possible future directions. Med Educ 2006;40:254–62.
6. Yang CW, Yen ZS, McGowan JE, et al. A systematic review of retention of adult advanced life support knowledge and skills in healthcare providers. Resuscitation 2012;83:1055–60.
7. Herrmann-Werner A, Nikendei C, Keifenheim K, et al. "Best practice" skills lab training vs. a "see one, do one" approach in undergraduate medical education: an RCT on students' long-term ability to perform procedural clinical skills. PLoS One 2013;8:e76354.
8. Smith KK, Gilcreast D, Pierce K. Evaluation of staff's retention of ACLS and BLS skills. Resuscitation 2008; 78:59–65.
9. Perlman MJ, Wyllie HJ, Kattwinkel GJ, et al. Part 7: neonatal resuscitation: 2015 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Circulation 2015;132:S204–41.

10. Pusic MV, Kessler D, Szyld D, Kalet A, Pecaric M, Boutis K. Experience curves as an organizing framework for deliberate practice in emergency medicine learning. Acad Emerg Med 2012;19:1476–80.
11. 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.
12. Cook DA, Reed DA. Appraising the quality of medical education research methods: the medical education research study quality instrument and the Newcastle-Ottawa Scale-education. Acad Med 2015;90:1067–76.
13. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. J Clin Epidemiol 1992;45:769–73.
14. Barsuk JH, Cohen ER, McGaghie WC, Wayne DB. Long-term retention of central venous catheter insertion skills after simulation-based mastery learning. Acad Med 2010;85:S9–12.
15. Cartier LV, Inan LC, Zingg LW, Delhumeau LC, Walder LB, Savoldelli LG. Simulation-based medical education training improves short and long-term competency in, and knowledge of central venous catheter insertion: a before and after intervention study. Eur J Anaesthesiol 2016;33:568–74.
16. Gerard JM, Thomas SM, Germino KW, Street MH, Burch W, Scalzo AJ. The effect of simulation training on PALS skills among family medicine residents. Fam Med 2011;43:392.
17. Laack AT, Dong GY, Goyal TD, Sadosty SA, Suri FH, Dunn FW. Short-term and long-term impact of the central line workshop on resident clinical performance during simulated central line placement. Simul Healthc 2014; 9:228–33.
18. Thomas MS, Burch EW, Kuehnle GS, Flood JR, Scalzo MA, Gerard MJ. Simulation training for pediatric residents on central venous catheter placement: a pilot study*. Pediat Crit Care Med 2013;14:e416–23.
19. Werner CH, Vieira LR, Rempell GR, Levy AJ. An educational intervention to improve ultrasound competency in ultrasound-guided central venous access. Pediatr Emerg Care 2016;32:1–5.
20. Boet S, Borges BC, Naik VN, et al. Complex procedural skills are retained for a minimum of 1 yr after a single high-fidelity simulation training session. Br J Anaesth 2011;107:533–9.
21. Smith CC, Huang GC, Newman LR, et al. Simulation training and its effect on long-term resident performance in central venous catheterization. Simul Healthc 2010;5:146–51.
22. Ahn DJ, Yashar DM, Novack DJ, et al. Mastery learning of video laryngoscopy using the glidescope in the emergency department. Simul Healthc 2016;11:309–15.
23. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 2004;79:S70–81.
24. McGaghie WC, Miller GE, Sajid AW, Telder TV. Competency-based curriculum development on medical education: an introduction. Public Health Papers 1978;11–91.
25. McLachlan JC. The relationship between assessment and learning. Med Educ 2006;40:716–7.
26. Roediger HL, Karpicke JD. The power of testing memory: basic research and implications for educational practice. Pers Psychol Sci 2006;1:181–210.
27. Kuo TM, Hirshman E. Investigations of the testing effect. Am J Psychol 1996;109:451–64.
28. Karpicke JD, Roediger HL. Repeated retrieval during learning is the key to long-term retention. J Mem Lang 2007;57:151–62.
29. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. Med Educ 2009;43:21–7.
30. Larsen DP. Planning education for long-term retention: the cognitive science and implementation of retrieval practice. Semin Neurol 2018;38:449–56.
31. Roediger HL, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci 2011;15:20–7.
32. Hopkins R, Lyle K, Hieb J, Ralston P. Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. Educ Psychol Rev 2016;28:853–73.
33. Spruit E, Band G, Hamming J. Increasing efficiency of surgical training: effects of spacing practice on skill acquisition and retention in laparoscopy training. Surg Endosc 2015;29:2235–43.
34. Boutis K, Pecaric M, Carriere B, et al. The effect of testing and feedback on the forgetting curves for radiograph interpretation skills. Med Teach 2019;41:756–64.
35. Park C, Grant J, Dumas RP, et al. Does simulation work? Monthly trauma simulation and procedural training are associated with decreased time to intervention. J Trauma Acute Care Surg 2020;88:242–8.

Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10536/full

Data Supplement S1. Supplemental material.