Available online at www.sciencedirect.com

Resuscitation Plus

journal homepage: www.elsevier.com/locate/resuscitation-plus

Review

Manikin physical realism for resuscitation education: A systematic review

Aaron Donoghue[a,*], Katherine Allan[b], Sebastian Schnaubelt[c], Andrea Cortegiani[d,e], Robert Greif[f,g], Adam Cheng[h], Andrew Lockey[i]

Abstract

Aim: To evaluate the impact of higher physical realism of manikins on educational and clinical outcomes during life support education.

Methods: This systematic review was conducted as part of the continuous evidence evaluation process of the International Liaison Committee on Resuscitation (ILCOR). A search of PubMed, Embase, and Cochrane was conducted from January 1, 2005 until April 30, 2024. Studies comparing training with higher physical realism manikins and lower realism manikins were eligible for inclusion. Studies comparing manikins to other forms of training (e.g. screen-based, virtual reality) were excluded. Risk of bias was assessed using Cochrane Risk of Bias 2 (RoB 2) for randomized trials and Risk Of Bias In Non-Randomized Studies of Interventions (ROBINS-I) for observational studies. For outcomes reported by four or more randomized studies, random effects *meta*-analysis using standardized mean difference was performed.

Results: Of the 1276 articles identified and screened, 21 articles comprised the final review (19 randomized trials, 2 observational studies). Meta-analysis of eight RCTs reporting simulation skill performance in a simulated clinical scenario at course conclusion demonstrated a benefit from the use of higher- realism manikins compared with lower realism manikins (standardized mean difference 0.66, 95% CI 0.08 – 1.25). Meta-analysis of seven RCTs reporting knowledge at course conclusion showed no significant difference between the use of both types of manikins. Significant risk of bias and a high degree of heterogeneity were found among the included studies.

Conclusion: This systematic review found that higher manikin realism during resuscitation training was associated with improved simulated clinical scenario performance at course conclusion; without an effect on knowledge at course conclusion. Future studies should examine the impact of resource requirements for high realism simulation on generalizability and implementation.

Keywords: Life support education, Physical realism, Manikin fidelity, Cardiopulmonary resuscitation

Introduction

Effective education of both laypersons and healthcare professionals in basic and advanced life support is one of the key components in improving survival from cardiac arrest.[1] Published studies have examined many novel approaches to instructional design, course delivery, or technology use to improve the effectiveness and durability of life support education.[2] The use of manikins with a greater degree of physical realism (so-called 'high-fidelity' manikins) has been implemented over the past three decades in the hopes of improving learner engagement by allowing them to 'suspend disbelief' to a greater extent and thus participate in life support training in a manner that more closely reflects their actual clinical performance.

An International Liaison Committee on Resuscitation (ILCOR) systematic review by Cheng et al in 2015 examined this research question, and concluded that higher manikin realism was associated with improved clinical performance at course conclusion.[3] This review discussed the inherent difficulty in balancing the cost and resource use associated with this technology, as well as the ongoing challenge of ensuring that the features demonstrated by a manikin with higher physical realism were incorporated into instructional design that appropriately addressed the needs of learners in a given setting. With recent advancements in simulation, new features of manikins, and updated educational approaches in teaching life

* Corresponding author at: Division of Critical Care Medicine, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.
E-mail address: donoghue@chop.edu (A. Donoghue).

https://doi.org/10.1016/j.resplu.2025.100940
Received 7 February 2025; Received in revised form 5 March 2025; Accepted 6 March 2025

support, the International Liaison Committee on Resuscitation (ILCOR) Education, Implementation and Teams Task Force (EIT) decided to perform this updated systematic review of contemporaneous evidence. We evaluated outcomes of life-support education studies comparing training with manikins with higher physical realism to training with manikins with lower physical realism.

Methods

This systematic review was part of the continuous evidence evaluation process by the International Liaison Committee on Resuscitation (ILCOR) Education, Implementation and Teams Task Force (EIT) and it was registered with PROSPERO (registration number CRD42024535044). The PROSPERO registration was subsequently updated with new wording for the intervention of interest ('physical realism' instead of 'fidelity') and the subset analysis of laypeople as learners was eliminated.

The systematic review was driven by a research question using the PICOST (Population, Intervention, Control, Outcomes, Study Design, and Timeframe) format: In learners training in any form of life support (P), does instruction using manikins with a greater degree of physical realism (I), compared to instruction with a lesser degree of physical realism (C), affect educational outcomes (skill, knowledge, attitudes) or clinical processes (change in healthcare provider behavior, patient outcomes, cost or resource utilization) (O)? Randomized controlled trials and non-randomized studies (non-randomized controlled trials, interrupted time series, controlled before-and-after studies, cohort studies) were included in any language as long as there was an English abstract available. Unpublished studies (e.g., conference abstracts, trial protocols) were excluded (S). Databases were searched over a timeframe from January 1, 2005 to April 30, 2024 (T); the 2005 start date was chosen based on the dates of published studies included in the previous systematic review being from 2005 onward.

Physical realism was defined as the presence of physical features (visible, palpable, audible) designed to increase the resemblance of the manikin to an actual patient. Eligible studies compared outcomes between groups using manikins of two different levels of realism; this was most commonly a static manikin compared with a higher degree of physical realism simulator, but no specific levels of realism were determined *a priori* to be a prerequisite for eligibility. Studies that compared manikins of any level of realism to other content delivery modalities (e.g. screen-based, virtual reality) were excluded.

Search strategy

We searched EMBASE.com, Medline, and Cochrane and citations of included studies were subsequently reviewed for additional eligible studies (Appendix).

Study selection and data collection

The titles and abstracts of all potentially eligible studies were screened for inclusion by pairs of independent reviewers. The full text of included studies were checked against the inclusion and exclusion criteria independently in these pairs of reviewers. Any disagreements between the reviewers at either stage were resolved by discussion finding consensus. Data from each study were independently extracted by a pair of reviewers and grouped separately according to the predefined outcomes.

Risk of bias assessment

Risk of bias was assessed by the same pairs of reviewers. The Cochrane Risk of Bias Tool for randomized trials (RoB 2) was used for randomized studies[4]; the Risk of Bias In Non-randomised Studies of Interventions (ROBINS-I) for observational studies.[5]

Data analysis and synthesis

Two reviewers independently assessed the certainty of evidence of each outcome using the GRADE approach (Grades of Recommendation, Assessment, Development and Evaluation). In case of disagreement, consensus was reached by discussion. We conducted both quantitative and qualitative syntheses of the evidence. For specific outcomes reported by more than four studies with sufficient data (point estimate and estimate of variance), we performed random-effects *meta*-analysis to quantitatively pool results using a standardized mean difference (SMD) to allow direct comparison. In interpreting the clinical significance of our results, we emphasized confidence intervals (CI) in relation to Cohen's effect size (or SMD) classifications, where SMD > 0.8 large; SMD 0.5–0.8 = moderate; SMD 0.2–0.5 = small; and SMD < 0.2 = negligible.[6] Data was entered into Review Manager (RevMan5, The Cochrane Collaboration, Oxford, UK) to calculate SMD, 95% confidence intervals and statistical heterogeneity. We quantified between-study inconsistency using the I^2 statistic, which estimates the percentage of variability not due to chance. I^2 values > 50% indicate large inconsistency or heterogeneity. Studies that could not be combined in the above quantitative synthesis of results were analysed and summarized in a narrative fashion.

Results

The search identified 1250 articles after duplicates were removed. Of these, 1217 articles were excluded during title and abstract screening, leaving 33 full-text articles to be screened for eligibility. During full-text review, 12 articles were subsequently excluded, leaving 21 articles comprising the final review.[7–27] The PRISMA diagram for the review is shown in Fig. 1. Fourteen of the included studies were among the ones included in the previously published systematic review[3]; seven additional studies are included in the present review. All studies involved healthcare professionals and/or healthcare trainees. 15 studies were performed in North America[8–16,18,22,24–27]; 4 in Asia[7,19,21,23]; one in Europe[17]; and one in Australia.[20] 12 studies used adult scenarios[7,10,14–18,20–24]; 4 used pediatric scenarios[9,12,25,27]; 4 used neonatal scenarios.[8,11,13,19] Table 1 contains the characteristics of these included studies.

Risk of bias assessments are shown in Table 2; all studies were found to have at least some concern for bias. Thirteen studies (twelve RCTs and one observational study) had previous assessment of bias when they were included in the 2015 systematic review[3]; the risk of bias assessment of the eight additional included studies are reported in Table 2.

The GRADE summary of evidence is shown in Table 3. Using GRADE criteria, evidence for each of the outcomes was found to be of low or very low certainty.

Skill performance

Eight RCTs with 279 intervention subjects and 271 controls reported skill at course conclusion assessed by overall clinical performance in a simulated scenario.[9,10,12–14,16,19,24] Four RCTs assessed perfor-

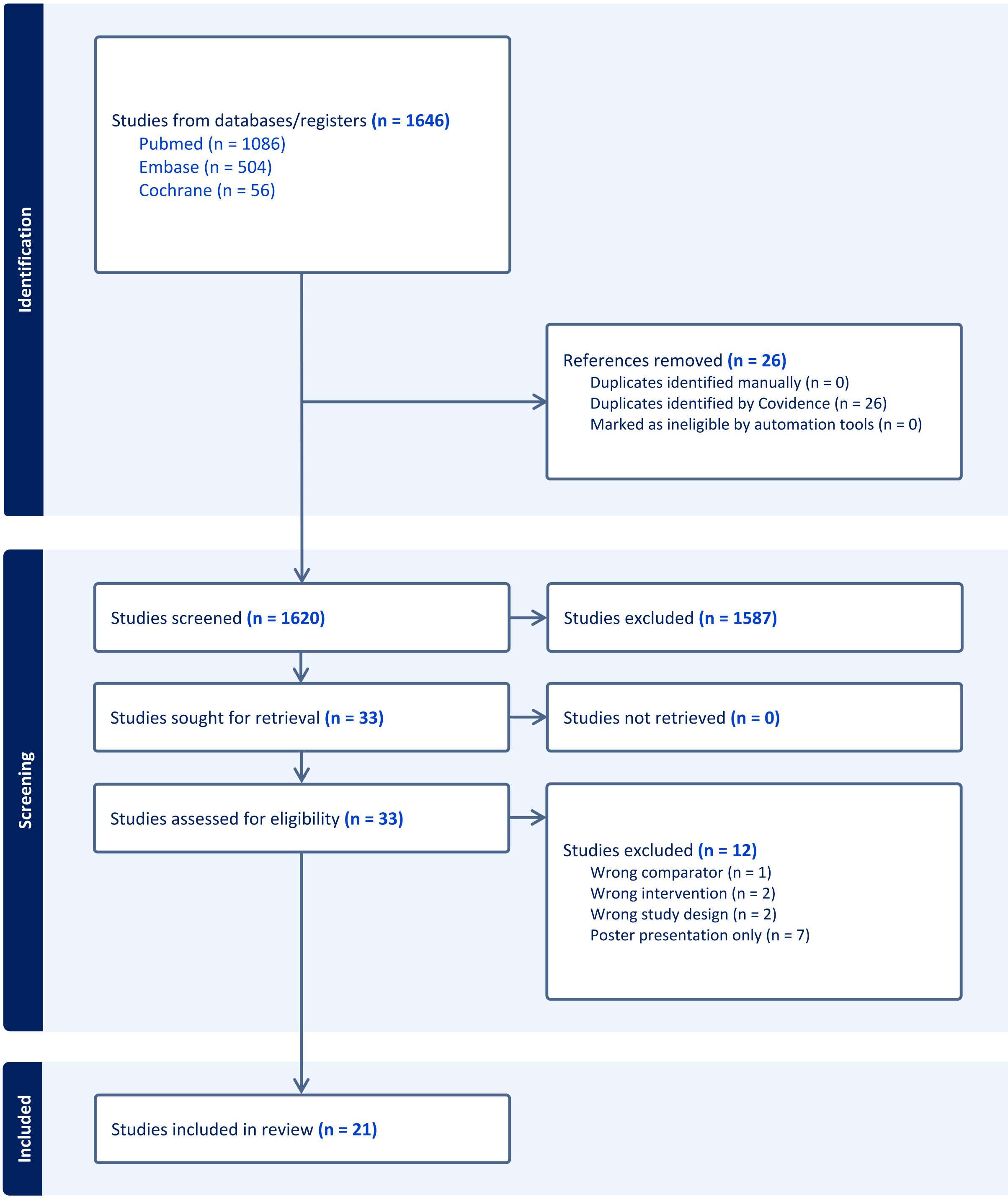

Fig. 1 – PRISMA Diagram.

mance in an adult scenario[10,14,16,24]; two in a pediatric scenario[9,12]; and two in a neonatal scenario. [13,19] Two studies used the NRP Megacode checklist as the outcome of interest[13,19]; two used the ACLS Megacode checklist[14,16]; the remaining four studies used instruments created or modified for their specific studies. [9,10,12,24] Four of the eight studies included an explicit description of rater training for use of the outcome instruments in the study. [9,13,16,19] On *meta*-analysis, a significant effect favoring manikins with higher realism was found (standard mean difference 0.66, 95% CI 0.08 – 1.25). A high degree of heterogeneity among studies was found (I^2 = 90%). (Fig. 2).

In addition to the eight RCTs included in the above *meta*-analysis, two additional RCTs with 51 intervention study participants and 56 controls compared skill at course conclusion but did not report sufficient measures of variance for inclusion in the *meta*-analysis. Both RCTs found no difference in skill performance at course completion. [11,20] Four RCTs measured and reported skill performance before and after training in addition to comparing low- and higher-realism

Table 1 – Data extraction table.

Study, year	Study Design	Subjects and Setting	Number of participants	Content	Outcome measures	Results
Aquel, 2014	RCT	Nursing students, Jordan	90 (45 per group)	BLS	K (pre-post test); S (CPR post test); both outcomes at completion and 3 months	Course conclusion CPR skills: CG 11.58 (1.63) vs IG 13.13 (1.01) ($p < 0.001$) Knowledge: CG 11.22 (0.90) vs IG 12.67 (1.07) $p < 0.001$). 3 months: CPR Skills: CG 10.31(1.88) vs IG 12.80 (1.44) ($p < 0.001$); Knowledge: CG 10.07(1.44) vs IG 12.27 (1.14)$p < 0.001$)
Campbell, 2009	RCT	1st year residents, Canada	15 (IG 8, CG 7)	NRP, 2 scenarios	S (number of redirections from facilitator; time to task completion; time to intubation)	Fewer redirections from facilitator in IG; no difference in time to intubate or other tasks
Cheng, 2013	RCT (factorial design)	Resident teams, USA	IG 45 teams; CG 45 teams	Single PALS scenario	S (clinical performance (CPT); teamwork (BAT)); K (MCQ test)	No significant differences
Conlon, 2014	RCT	EM residents, USA	54 (IG 18; CG 36 (18 in lo-fi, 18 in mid-fi))	ACLS	S (clinical performance); K (written exam)	No significant differences
Curran, 2015	RCT	Medical students, Canada	IG 31; CG 35	NRP	S (clinical, teamwork)	No significant differences
Donoghue, 2009	RCT	Residents, USA	IG 25; CG 26	PALS scenarios	S (clinical performance)	Greater skill improvement in IG ($p = 0.007$)
Finan, 2012	RCT	Neonatal fellows, Canada	16 (crossover)	NRP scenarios	S (megacode performance; ANTS score)	No significant differences
Hoadley, 2009	RCT	Allied heath professionals, USA	IG 29; CG 24	ACLS	S (clinical scenarios); K (Written test)	No significant differences
King, 2011	RCT	Nursing students, USA	49 (IG 24, CG 25)	ACLS	K (written test); S scenario checklists at 2 wk and 2 mons	No significant differences
Lo, 2011	RCT	Medical students, USA	86 (IG 45, CG 41)	ACLS	S (megacode performance at completion and at 1 year)	At completion: IG > CG (83% vs 70%, $p < 0.001$); at one year NSD
Massoth 2019	RCT	Medical students, Germany	135 (IG 67, CG 68)	ALS	K: written test; S: practical performance assessed by video analysis; A: self-assessment, confidence	Knowledge: No significant difference ($p > 0.05$). Skill: CG superior in breathing control ($p = 0.012$), continous chest compression while charging the defibrillator ($p = 0.017$), electrocardiogram analysis ($p = 0.021$), time interval between electric shocks ($p = 0.016$).
McCoy 2018	RCT	Medical students, USA	70 (IG 35, CG 35)	BLS	S: compression rate, depth, recoil, compression fraction; time to EMS activation	Skill: mean compression depth IG 4.57 cm vs CG 3.89 cm ($p = 0.02$) Compression fraction IG 72% vs CG 70% ($p = 0.01$) Mean compression rate IG 123.3 cpm vs CG 116 cpm ($p = 0.06$) Full recoil: NSD Time to activation of EMS: IG 24.7 sec vas CG 79.5 s ($p = 0.007$)

Table 1 (continued)

Study, year	Study Design	Subjects and Setting	Number of participants	Content	Outcome measures	Results
Nimbalkar, 2015	RCT	Medical students, India	101 (IG 51, CG 53)	NRP	S: NRP megacode scenario; K: NRP written test (both at course completion and 3 mos)	No significant differences
Owen, 2006	RCT	Residents, Australia	41 (IG 20, CG 21)	Adult cardiac arrest scenario	S: clinical performance score; K: written test	K: IG > CG (p = 0.026); S: NSD
Rishipathak 2020	RCT	EMS students, India	100 (unsure of group size)	ACLS	A (learner satisfaction and confidence)	Higher confidence in IG; higher preference (2 out of 5 items) in IG
Roh 2014	Observational study	nursing students, South Korea	163 (IG 28, CG 135)	Adult cardiac arrest scenario	A: self-efficacy scale	Greater improvement in self-efficacy scores in IG (p < 0.001)
Rogers, 2009	Observational study	Nursing students, USA	34 (IG 16; CG 18)	ACLS	S: scenario performance; K: written test	S: higher scenario score in IG (p = 0.01); K: greater pre- to post- test improvement (p = 0.002)
Settles, 2011	RCT	Allied health professionals, USA	146 (IG 73, CG 73)	ACLS	S: scenario performance (3–6 mos, 6–9 mos); K: written test	No significant differences
Stellflug, 2018	RCT	Healthcare providers, USA	94 (IG 28; CG 66)	PALS	S: time to task performance; Knowledge: PALS written exam	Skill: NSD; Knowledge: NSD btw groups at course completion; improved knowledge retention at 6 months post course in IG group (exam score 27.7 vs. 26.2, p = 0.042)
Thomas, 2010	RCT	Residents, USA	62 (IG 31, CG 31)	NRP	S: NRP megacode score	No significant differences
Tufts, 2021	RCT	Medical students, USA	50 (IG 27, CG 23)	PALS scenario	S: resuscitation task completion; A: self-reported confidence	Skill: IG > CG in 4 of 11 specific tasks (p < 0.001 for all); self confidence IG > CG in 3 of 14 domains

Abbreviations: RCT: randomized controlled trial; S: skill; K: knowledge; A: attitude; IG: intervention group (higher fidelity); CG: control group (lower fidelity); BLS: Basic life support; CPR: cardiopulmonary resuscitation; NRP: Neonatal Resuscitation Program; PALS: Pediatric advanced life support; ACLS: Advanced cardiac life support; EM: emergency medicine; NSD: no significant difference; PPV: positive pressure ventilation; MCQ: multiple choice questions.

Table 2 – Assessment of bias tables for studies from 2015 to present.

RCTs (RoB 2)

1st Author	Year	Randomization	Deviations from interventions	Missing data	Outcome measurement	Outcome reporting	Overall
Aqel	2014	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	***Some concerns***
Massoth	2019	Low	Low	Low	Some concerns	Some concerns	***Some concerns***
McCoy	2018	Low	Low	Low	Some concerns	Some concerns	***Some concerns***
Nimbalkar	2015	Low	Low	Low	Some concerns	Low	***Some concerns***
Rishipithak	2020	Some concerns	Some concerns	High	High	Some concerns	***High***
Stellflug	2018	Some concerns	Low	Low	Some concerns	Low	***Some concerns***
Tufts	2021	Some concerns	Low	Low	Some concerns	Some concerns	***Some concerns***

Observational studies (ROBINS-I)

1st Author	Year	Confounding	Participant selection	Intervention classification	Deviations from interventions	Missing data	Outcome measurement	Outcome reporting	Overall
Roh	2022	Serious	Low	Low	Low	Serious	Serious	Low	***Serious***

manikin groups. [9,12,19,20] All four studies demonstrated improvement in performance post-training when compared with pre-training irrespective of the level of manikin realism.

Three RCTs with 71 intervention study participants and 108 controls examined time to critical task completion at course conclusion.[8,18,25] Two RCTs found faster time-to-task completion (EMS activation) [18] and intervention/reassessment cycle in simulations with higher realism manikins. [25] One RCT found no difference in time to intubation during NRP training.[8]

Three RCTs with 96 intervention study participants and 97 controls examined teamwork performance at course conclusion; all three used a separate outcome instrument on which raters were specifically trained.[9,20,26] One RCT found improved teamwork behaviors following training with higher realism manikins compared with lower realism; one RCT found improved teamwork behavior during one of three assessment scenarios following training with higher realism[20]; one RCT found no difference in teamwork scores between groups.[9]

Three RCTs with 163 intervention subjects and 159 controls examined clinical performance at 3 months or longer following course conclusion. [7,16,24] One RCT in nursing students found better clinical performance in a CPR scenario 3 months after training with higher realism manikins[7]; two studies of advanced cardiac life support (ACLS) skills found no difference at 3 months post training or at one year post training.[16,24]

Two RCTs with 80 intervention study participants and 80 controls examined specific CPR parameters at course conclusion. One RCT found a greater degree of improvement as measured at course completion by the American Heart Association CPR skills checklist among study participants trained on higher realism manikins[7]; one RCT found better compression depth and compression fraction immediately post training among study participants trained on higher realism manikins.[18]

Knowledge acquisition and retention

Seven RCTs with 511 intervention study participants and 505 controls reported on knowledge at course conclusion as measured by a written test. [7,9,10,14,17,19,24] Five RCTs assessed performance in an adult scenario[7,10,14,17,24]; one in a pediatric scenario,[9] and one in a neonatal scenario.[19] On *meta*-analysis, no effect favoring higher realism manikins was found (standard mean difference 0.24, 95% CI −0.23 – 0.71). Significant heterogeneity was again found amongst these studies ($I^2 = 92\%$).(Fig. 3).

In addition to the seven RCTs included in the above *meta*-analysis, three additional RCTs with 72 intervention study participants and 112 controls and one observational study of 34 study participants compared knowledge at course conclusion but did not report sufficient measures of variance for inclusion in metaanalysis.[15,20,22,25] One RCT found improved knowledge at course completion[20]; two RCTs and one observational study found no difference.[15,22,25]

Three RCTs with 146 intervention study participants and 184 controls reported on knowledge assessed at 3 months or longer after course conclusion. [7,24,25] Two RCTs found improved knowledge following higher realism manikin training (3 months after BLS training[7]; 6 months after PALS training[25]); one RCT found no difference in ACLS knowledge at 6–9 months post training.[24]

Attitudes

Ten RCTs with 347 intervention study participants and 471 controls reported on learner preference and/or confidence at course conclusion.[8,10,11,13,14,17,21,23,24,27] Seven RCTs found study participants reporting higher effectiveness of training with higher realism manikins[8,10,11,17,21,23,27]; three RCTs found no difference.[13,14,24]

Clinical outcomes, costs and resources

No studies included in the review examined care delivery to patients, patient clinical outcomes, cost efficiency, or resource utilization.

Discussion

This systematic review provides low to very low certainty evidence that using manikins with higher degrees of physical realism during life support training may lead to improved educational outcomes. In particular, clinical skill performance in a simulated scenario at course conclusion was found to be significantly improved in learners trained on higher realism manikins. We found no effect of manikin realism on knowledge at course completion. Our findings support and strengthen the recommendations yielded by the 2015 systematic review by Cheng et al. examining this same research question.[3]

Table 3 – GRADE Summary of Evidence.

Question: Higher fidelity manikins compared to lower fidelity manikins for life support education
Setting: Life support education settingsBibliography:

Certainty assessment							of patients		Effect		Certainty	Importance
of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	higher fidelity manikins	lower fidelity manikins	Relative (95% CI)	Absolute (95% CI)		
Skill –- megacode/scenario performance at course conclusion												
8[9,10,12–14,16,19,24]	randomised trials	serious[a]	serious[b]	not serious	serious[b]	none	279	271	–	SMD **0.66 SD higher** (0.08 higher to 1.25 higher)	⊕○○○ Very low[a,b]	CRITICAL
Knowledge –- written / MCQ test at course conclusion												
7[7,9,10,14,17,19,24]	randomised trials	serious[a]	serious[b]	serious[c]	serious[b]	none	511	505	–	SMD **0.24 SD higher** (0.23 lower to 0.71 higher)	⊕○○○ Very low[a,b,c]	CRITICAL
Time to clinical task completion (course conclusion)												
3[8,18,25]	randomised trials	serious[a]	serious[d]	not serious	not serious	none	71 intervention and 108 control subjects in 3 RCTs examined time to task completion. Two RCTs found faster time-to-task completion (EMS activation (McCoy) and intervention/reassessment cycle (Stellflug)). One RCT (Campbell) found no difference in time to intubation during NRP training.				⊕⊕○○ Low[a,d]	IMPORTANT
Skill –- teamwork (course conclusion)												
3[9,20,26]	randomised trials	serious[a]	serious[b]	not serious	serious[b]	none	96 intervention / 97 control subjects in 3 RCTs examined teamwork with different levels of manikin fidelity. One RCT (Thomas) found improved teamwork behaviors following training with higher fidelity manikins compared with lower fidelity; one RCT (Owen) found improved teamwork behavior during one of three assessment scenarios following training with higher fidelity; one RCT (Cheng) found no difference in teamwork scores between groups.				⊕○○○ Very low[a,b]	IMPORTANT
Skill –- clinical performance at 3 months or greater												
3[7,16,24]	randomised trials	serious[a]	serious[b]	not serious	serious[b]	none	163 intervention / 159 control subjects in 3 RCTs examined clinical skill (scenario-based assessment) at 3 months or greater after training. One RCT (Aqel) in nursing students found better clinical performance in a CPR scenario 3 months after training with higher fidelity manikins; two studies of ACLS skills found no difference at 3 months post training (Lo) or at one year post training (Settles).				⊕○○○ Very low[a,b]	IMPORTANT
Knowledge –- written/MCQ test at 3 to 6 months												
3[7,24,25]	randomised trials	serious[a]	serious[b]	serious[c]	serious[b]	none	146 intervention / 184 control subjects in 3 RCTs examined knowledge (written test assessment) at 3 months or greater following training. Two RCTs found improved knowledge following higher fidelity manikin training (3 months after BLS training (Aqel); 6 months after PALS training (Stellflug)); one RCT found no difference in ACLS knowledge at 6–9 months post training (Settles).				⊕○○○ Very low[a,b,c]	IMPORTANT
Skill –- CPR parameters at course conclusion												
2[7,18]	randomised trials	serious[a]	serious[b]	not serious	serious[b]	none	2 RCTs with 80 intervention subjects and 80 controls. One RCT found a greater degree of improvement as measured at course completion by the American Heart Association CPR skills checklist among subjects trained on higher fidelity manikins (Aqel); one RCT found better compression depth and compression fraction immediately post training among subjects trained on higher fidelity manikins. (McCoy)				⊕○○○ Very low[a,b]	IMPORTANT
Affective responses												
10[8,10,11,13,14,17,21,23,24,27]	randomised trials	very serious[a,d]	serious[b]	serious[c]	very serious[e]	none	10 RCTs with 347 intervention subjects and 471 controls. Seven RCTs found subjects reporting higher effectiveness of training with higher fidelity manikins (Campbell, Conlon, Curran, Massoth, Rishipathak, Roh, Tufts); three RCTs found no difference. (Finan, Hoadley, Settles)				⊕○○○ Very low[a,b,c,d,e]	IMPORTANT

CI: confidence interval; **SMD:** standardised mean difference.

Explanations.

[a] Blinding of participants not possible in RCTs.
[b] Different assessment tools across multiple studies.
[c] No proven correlation between written exam scores and clinical performance.
[d] Different tasks being assessed across studies.
[e] Data from survey instruments measuring different constructs.

Study or Subgroup	Higher fidelity Mean	SD	Total	Mean	SD	Total	Weight	Std. mean difference IV, Random, 95% CI
Cheng 2013	7.9	7	23	6.3	7.1	22	12.5%	0.22 [-0.36 , 0.81]
Conlon 2014	66.3	9	18	60.1	8.9	18	12.0%	0.68 [0.00 , 1.35]
Donoghue 2009	11.1	1.8	25	4.8	1.7	26	10.6%	3.55 [2.64 , 4.45]
Finan 2012	72.2	9	16	78.2	11.7	16	11.8%	-0.56 [-1.27 , 0.15]
Hoadley 2009	26.1	2.5	29	24.7	3.7	24	12.7%	0.45 [-0.10 , 0.99]
Lo 2011	41.5	5.6	45	35	6	41	13.2%	1.11 [0.66 , 1.57]
Nimbalkar 2015	31.5	6.1	50	29.7	5.6	51	13.4%	0.31 [-0.09 , 0.70]
Settles 2011	17.2	1.4	73	17.2	1.6	73	13.7%	0.00 [-0.32 , 0.32]
Total (95% CI)			**279**			**271**	**100.0%**	**0.66 [0.08 , 1.25]**

Heterogeneity: Tau² = 0.61; Chi² = 70.00, df = 7 (P < 0.00001); I² = 90%
Test for overall effect: Z = 2.24 (P = 0.02)
Test for subgroup differences: Not applicable

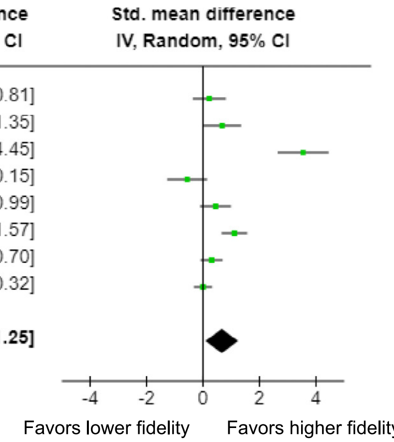

Fig. 2 – Meta-analysis of RCTs reporting scenario-based clinical skill performance at course conclusion.

Study or Subgroup	Higher fidelity Mean	SD	Total	Lower fidelity Mean	SD	Total	Weight	Std. mean difference IV, Random, 95% CI
Aqel 2014	12.7	1.1	45	11.2	0.9	45	14.0%	1.48 [1.01 , 1.95]
Cheng 2013	4	1.8	228	4.9	1.8	224	15.8%	-0.50 [-0.69 , -0.31]
Conlon 2014	40.8	14	18	41.3	14.1	18	12.4%	-0.03 [-0.69 , 0.62]
Hoadley 2009	90.3	7.8	29	87.7	9.3	24	13.3%	0.30 [-0.24 , 0.85]
Massoth 2019	16.5	2	67	16.3	2.6	68	14.9%	0.09 [-0.25 , 0.42]
Nimbalkar 2015	35.2	2.7	51	33.9	3.3	53	14.6%	0.43 [0.04 , 0.82]
Settles 2011	23.6	1.4	73	23.6	1.4	73	15.0%	0.00 [-0.32 , 0.32]
Total (95% CI)			**511**			**505**	**100.0%**	**0.24 [-0.23 , 0.71]**

Heterogeneity: Tau² = 0.36; Chi² = 71.13, df = 6 (P < 0.00001); I² = 92%
Test for overall effect: Z = 0.99 (P = 0.32)
Test for subgroup differences: Not applicable

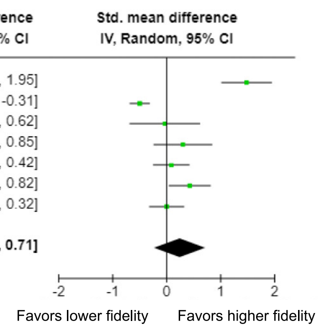

Fig. 3 – Meta-analysis of RCTs reporting knowledge assessment at course conclusion.

All of the included studies had significant risk of bias, and large heterogeneity was found among pooled studies in *meta*-analyses. Additionally, significant variability existed between studies with respect to instructional content, learner group backgrounds and experience levels, and the types of assessment method(s) used. Aside from the outcomes subjected to *meta*-analysis, the other outcomes (e.g. CPR parameters, task performance) were reported only by three or fewer studies; even in these small subsets, there were differences in the study population and life support course being studied. All of these issues make drawing strong conclusions about the effect of manikin realism difficult. A total of six studies reported either skill or knowledge at remote time points (as opposed to at course conclusion); this subset of studies showed a similar degree of variability, making conclusions about the durability of learning and skill and/or knowledge decay impossible.

The recommendations made by ILCOR in 2020 for use of higher realism manikins stipulate that they should be used by centers that have sufficient resources to implement them in an optimal way.[28,29] This includes financial resources for the purchase and maintenance of such simulators; infrastructure such as space and information technology; and personnel with sufficient training and expertise. Importantly, none of the included studies in this review directly addressed the need for financial, infrastructural, and personnel resources associated with higher realism manikin use. It is not possible to extrapolate these recommendations to settings lacking these resources. A recent systematic review by Issa et al. examined the impact of manikin realism across all medical education areas in lower and middle income countries and found no significant benefit in the used of higher realism training models, leading to a recommendation that lower realism models be used based on the associated resource requirements.[30] While Issa and colleagues did not focus exclusively on life support education, it is possible that the cost necessary to use such technology outweighs the educational benefit in such settings. Further studies should more closely examine resource requirements for resuscitation training with more realistic manikins and determine which manikin features are the most contributory for learning for the most learners.

Limitations, knowledge gaps, and future research

Several limitations to the findings of this review should be acknowledged. As described above, the studies included in this review demonstrated marked heterogeneity, making summative conclusions about the impact of manikin realism on educational outcomes difficult. Data with sufficient detail for *meta*-analysis was reported in eight RCTs for skill and seven RCTs for knowledge at course conclusion; however, given that these assorted studies were conducted using differing content (e.g. PALS cs. NRP vs. ALS) a significant risk of bias remains based on inconsistency. The application of unified assessment of the learning outcome in future studies is highly recommended to enable comparative studies and *meta*-analysis.

The results of this review highlight several important knowledge gaps with regard to manikin realism. The majority of included studies were performed at single institutions; determining the generalizability of these results would require studying the uniformity across institutions of the types of manikins used and the associated instructor training. Higher realism manikins are associated with significant cost; no included studies attempted to examine cost effectiveness or return on investment. Finally, as discussed above, the substantial resource requirements for implementing training with higher realism manikins, in terms of cost, personnel, and infrastructure, make it impossible to translate these findings to resource-limited settings.

Conclusions

This systematic review provides very low certainty evidence supporting the use of manikins with higher degrees of physical realism during life support education. In one *meta*-analysis, an association between higher manikin realism use and improved clinical performance at course conclusion was found. A high degree of risk of bias and heterogeneity among studies makes inference of a stronger association with learning outcomes difficult. Future studies should examine the impact of the resource requirements for higher realism manikin use as well as the features in such manikins that maximize functional task alignment with specific learner needs and objectives.

Conflicts of interest

Each of the authors are members of of ILCOR's task force on Education, Implementation, and Teams. RG is the ERC Director of Guidelines and ILCOR, and Chair of ILCOR's task force on Education, Implementation, and Teams. RG, AL and AC are Editorial Board members for *Resuscitation Plus*. AD and AC are first authors of two of the included studies in the review and did not participate in risk of bias assessment and GRADE rating, other members of the task force writing group reviewed and extracted these studies. AC and AL are co-authors of the previous ILCOR systematic review from which data on 12 of the included studies was extracted; two separate authors (AD and KA) were responsible for the assessment of quality of that review using A Measurement Tool to Assess Systematic Reviews (AMSTAR-2) prior to its inclusion.[31]

CRediT authorship contribution statement

Aaron Donoghue: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Katherine Allan:** Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation, Conceptualization. **Sebastian Schnaubelt:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Andrea Cortegiani:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Robert Greif:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration. **Adam Cheng:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Andrew Lockey:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Jennifer Lege-Matsuura, MSLIS, AHIP, and Elizabeth S. Blake, MLS, AHIP, both from the Holman Biotech Commons at the University of Pennsylvania, Philadelphia, PA, USA; and Julie Reed, MA, MLIS, from the Department of Continuing Education at the Children's Hospital of Philadelphia, for their support in developing, reviewing, and conducting the search strategies.

The following ILCOR EIT Taskforce members are acknowledged as collaborators on this systematic review: Cristian Abelairas-Gómez, Jan Breckwoldt, Kathryn Eastwood, Barbara Farquharson, Ming-Ju Hsieh, Tracy Kidd, Ying-Chih Ko, Kasper G. Lauridsen, Yiqun Lin, Tasuku Matsuyama, Sabine Nabecker, Kevin Nation, Alexander Olaussen, Chih-Wei Yang, and Joyce Yeung.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.resplu.2025.100940.

Author details

[a]*Departments of Critical Care Medicine and Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA* [b]*Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada* [c]*Department of Emergency Medicine, Medical University of Vienna, Austria* [d]*Department of Precision Medicine in Medical, Surgical and Critical Care (Me.Pre.C.C.), University of Palermo, Palermo, Italy* [e]*Department of Anesthesia, Analgesia, Intensive Care and Emergency, University Hospital Policlinico Paolo Giaccone, Palermo, Italy* [f]*University of Bern, Bern, Switzerland* [g]*School of Medicine, Sigmund Freud University Vienna, Vienna, Austria*[h]*KidSIM Simulation Program, Alberta Children's Hospital, University of Calgary, Calgary, Alberta, Canada* [i]*School of Human and Health Sciences, University of Huddersfield, Huddersfield, United Kingdom*

REFERENCES

1. Soreide E, Morrison L, Hillman K, et al. The formula for survival in resuscitation. Resuscitation 2013;84:1487–93. https://doi.org/10.1016/j.resuscitation.2013.07.020. 20130803.
2. Cheng A, Nadkarni VM, Mancini MB, et al. Resuscitation education science: educational strategies to improve outcomes from cardiac arrest: a scientific statement from the american heart association. *Circulation* 2018;138:e82–e122. https://doi.org/10.1161/CIR.0000000000000583.

3. Cheng A, Lockey A, Bhanji F, et al. The use of high-fidelity manikins for advanced life support training–A systematic review and meta-analysis. Resuscitation 2015;93:142–9. https://doi.org/10.1016/j.resuscitation.2015.04.004. 20150414.
4. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. https://doi.org/10.1136/bmj.l4898. 20190828.
5. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355(i4919). https://doi.org/10.1136/bmj.i4919. 20161012.
6. Cohen J. *Statistical power analysis for the behavioral sciences*. Routledge; 2013.
7. Aqel AA, Ahmad MM. High-fidelity simulation effects on CPR knowledge, skills, acquisition, and retention in nursing students. *Worldviews Evid Based Nurs* 2014;11:394–400. https://doi.org/10.1111/wvn.12063.
8. Campbell DM, Barozzino T, Farrugia M, et al. High-fidelity simulation in neonatal resuscitation. *Paediatr Child Health* 2009;14:19–23. https://doi.org/10.1093/pch/14.1.19.
9. Cheng A, Hunt EA, Donoghue A, et al. Examining pediatric resuscitation education using simulation and scripted debriefing: a multicenter randomized trial. *JAMA Pediatr* 2013;167:528–36. https://doi.org/10.1001/jamapediatrics.2013.1389.
10. Conlon LW, Rodgers DL, Shofer FS, et al. Impact of levels of simulation fidelity on training of interns in ACLS. *Hosp Pract (1995)* 2014;42:135–41. https://doi.org/10.3810/hp.2014.10.1150.
11. Curran V, Fleet L, White S, et al. A randomized controlled study of manikin simulator fidelity on neonatal resuscitation program learning outcomes. *Adv Health Sci Educ Theory Pract* 2015;20:205–18. https://doi.org/10.1007/s10459-014-9522-8.
12. Donoghue AJ, Durbin DR, Nadel FM, et al. Effect of high-fidelity simulation on Pediatric Advanced Life Support training in pediatric house staff: a randomized trial. *Pediatr Emerg Care* 2009;25:139–44. https://doi.org/10.1097/PEC.0b013e31819a7f90.
13. Finan E, Bismilla Z, Whyte HE, et al. High-fidelity simulator technology may not be superior to traditional low-fidelity equipment for neonatal resuscitation training. *J Perinatol* 2012;32:287–92. https://doi.org/10.1038/jp.2011.96.
14. Hoadley TA. Learning advanced cardiac life support: a comparison study of the effects of low- and high-fidelity simulation. *Nurs Educ Perspect* 2009;30:91–5.
15. King JM, Reising DL. Teaching advanced cardiac life support protocols: the effectiveness of static versus high-fidelity simulation. *Nurse Educ* 2011;36:62–5. https://doi.org/10.1097/NNE.0b013e31820b5012.
16. Lo BM, Devine AS, Evans DP, et al. Comparison of traditional versus high-fidelity simulation in the retention of ACLS knowledge. *Resuscitation* 2011;82:1440–3. https://doi.org/10.1016/j.resuscitation.2011.06.017.
17. Massoth C, Röder H, Ohlenburg H, et al. High-fidelity is not superior to low-fidelity simulation but leads to overconfidence in medical students. *BMC Med Educ* 2019;19:29. https://doi.org/10.1186/s12909-019-1464-7.
18. McCoy CE, Rahman A, Rendon JC, et al. Randomized controlled trial of simulation vs. standard training for teaching medical students high-quality cardiopulmonary resuscitation. *West J Emerg Med* 2019;20:15–22. https://doi.org/10.5811/westjem.2018.11.39040.
19. Nimbalkar A, Patel D, Kungwani A, et al. Randomized control trial of high fidelity vs low fidelity simulation for training undergraduate students in neonatal resuscitation. *BMC Res Notes* 2015;8:636. https://doi.org/10.1186/s13104-015-1623-9.
20. Owen DD, McGovern SK, Murray A, et al. Association of race and socioeconomic status with automatic external defibrillator training prevalence in the United States. *Resuscitation* 2018;127:100–4. https://doi.org/10.1016/j.resuscitation.2018.03.037.
21. Rishipathak P, Hinduja A, Sengupta N. A comparative analysis of self-efficacy in low fidelity vs high fidelity simulation post advanced cardiac life support (ACLS) sessions on cardiac arrest algorithm amongst EMS students of Pune, India. *Indian Journal of Public Health Research and Development* 2020;11:415–9. https://doi.org/10.37506/v11/i1/2020/ijphrd/193853.
22. Rodgers DL, Securro Jr S, Pauley RD. The effect of high-fidelity simulation on educational outcomes in an advanced cardiovascular life support course. *Simul Healthc* 2009;4:200–6. https://doi.org/10.1097/SIH.0b013e3181b1b877.
23. Roh YS. Effects of high-fidelity patient simulation on nursing students' resuscitation-specific self-efficacy. *Comput Inform Nurs* 2014;32:84–9. https://doi.org/10.1097/cin.0000000000000034.
24. Settles J, Jeffries PR, Smith TM, et al. Advanced cardiac life support instruction: do we know tomorrow what we know today?. *J Contin Educ Nurs* 2011;42(271–279)20110322. https://doi.org/10.3928/00220124-20110315-01.
25. Stellflug SM, Lowe NK. The effect of high fidelity simulators on knowledge retention and skill self efficacy in pediatric advanced life support courses in a rural state. *J Pediatr Nurs* 2018;39:21–6. https://doi.org/10.1016/j.pedn.2017.12.006.
26. Thomas EJ, Williams AL, Reichman EF, et al. Team training in the neonatal resuscitation program for interns: teamwork and quality of resuscitations. *Pediatrics* 2010;125:539–46. https://doi.org/10.1542/peds.2009-1635.
27. Tufts LM, Hensley CA, Frazier MD, et al. Utilizing high-fidelity simulators in improving trainee confidence and competency in code management. *Pediatr Qual Saf* 2021;6:e496.
28. Greif R, Bhanji F, Bigham BL, et al. Education, implementation, and teams: 2020 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Resuscitation 2020;156:A188–239. https://doi.org/10.1016/j.resuscitation.2020.09.014. 20201021.
29. Greif R, Bhanji F, Bigham BL, et al. Education, implementation, and teams: 2020 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Circulation 2020;142:S222–83. https://doi.org/10.1161/CIR.0000000000000896. 20201021.
30. Issa M, Furia F, Whaiba A, et al. Physical realism of simulation training for health care in low- and middle-income countries-a systematic review. *Simul Healthc* 2024;19:S41–9. https://doi.org/10.1097/SIH.0000000000000761.
31. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008. https://doi.org/10.1136/bmj.j4008. 20170921.