### Clinical question.

"For healthcare professionals (P), do simulation based learning methodologies (I), when compared to traditional (lectures) training (C), improve the acquisition of content knowledge, technical skills, and behavioral skills required for effective and safe resuscitation (O)?"

Is this question addressing an intervention/therapy, prognosis or diagnosis? Educational intervention

State if this is a proposed new topic or revision of existing worksheet: New topic

**Conflict of interest specific to this question**

Do any of the authors listed above have conflict of interest disclosures relevant to this worksheet? Yes, Khalid Aziz is an author on a randomized controlled trial that addresses the question in neonatal resuscitation.

**Search strategy (including electronic databases searched).**

PubMed search 9 Feb 2009: 966 hits (with 50 further hits up to 30 Sep 2009, with 18 relevant titles)


AND

((("Education/instrumentation"[Mesh] OR "Education/manpower"[Mesh] OR "Education/methods"[Mesh])) OR "education "[Subheading])) OR (("Computer Simulation"[Mesh]) OR "simulation")

For the remaining databases, the following strings were used in each search unless otherwise stated, based on the “method”, the “purpose”, and the “context”. Terms such as “clinical trial” and “randomized controlled trial” were avoided as they may not encompass educational environments. The initial search included all critical care environments, and not just “resuscitation” – subsequently, the search was narrowed to “resuscitation” in any field:

((SIMULATION) OR (SIMULATED) OR (IMMERSIVE) OR (COMPUTER ASSISTED INSTRUCTION) OR (VIDEORECORDING) OR (ROLE PLAYING) OR (MANIKINS))

AND

(("EDUCATION) OR (INSERVICE TRAINING) OR (EDUCATIONAL TECHNOLOGY) OR (TEACHING) OR (ATTITUDE) OR (CLINICAL COMPETENCE) OR (RETENTION))

AND

(("RESUSCITATION) OR (CRITICAL CARE) OR (CARDIOPULMONARY RESUSCITATION) OR (INTENSIVE CARE) OR (DELIVERY ROOMS) OR (PATIENT CARE TEAM))

Search was performed on 17 November 2008 as follows:

1443 MEDLINE hits downloaded to EndNotes

345 EMBASE hits downloaded to EndNotes (164 duplicates) (a further 60 hits up to 30 Sep 2009, with 18 relevant titles)

96 WILSON OMNIFILE hits to EndNotes (11 duplicates) (a further 9 hits up to 30 Sep 2009, with 1 relevant title)

740 SCOPUS hits (after limiting to medical, nursing or health professionals as >2000 hits) – to Endnotes (381 duplicates)

142 CINAHL hits when limited to RESUSCITATION and PERIODICALS – no new articles.

Overall the 17 November search (excluding CINAHL) revealed 2178 hits (excluding electronically detected duplicates and CINAHL). Limited the search to (RESUSCITATION in all fields) left 844 hits. Manual search to remove duplicates left 758 hits.


A search of the COCHRANE DATABASE OF SYSTEMATIC REVIEWS using “SIMULAT*” (88 hits, manually reviewed and none were relevant), and limited to “RESUSCITAT*” exposed 3 non-contributory reviews. An additional search of DARE and the COCHRANE CENTRAL REGISTER OF CONTROLLED TRIALS for “SIMULAT*” and “RESUSCITAT*” revealed 89 hits, but no new studies were found.

### State inclusion and exclusion criteria

**Inclusion criteria:**

Healthcare providers and trainees in healthcare provision

Studies based on immersive learning environments including simulation-based learning methodologies, including manikins, computer-based learning, and self-directed learning compared to a control of traditional methods or lower technical “fidelity”.

**Exclusion criteria:**

...
Studies with a control educational intervention
Randomized and non-randomized controls, including crossover studies.
Acquisition of content knowledge, technical skills, and behavioral skills as an outcome
Clinical outcomes
Resuscitation (all ages), including trauma, adult, pediatric, medical, and neonatal

**Exclusion criteria:**
Studies of the effect on lay providers: there were a large number of studies looking at whether lay providers could be trained using traditional vs other immersive environments.
Reviews, editorials, program descriptions, and case reports
Pretest-posttest designs with no control or historical controls: many studies observe the change in learner/team performance or patient outcome before and after an immersive learning intervention. These studies were excluded, as were studies where the intervention represented significantly more training in time than the control. All these studies can conclude is that more training (whether in immersive environments or not) might impact learner, practice or patient outcomes.
Use of simulations to compare different devices
Resuscitation studies that do not include the key steps of airway management, ventilation, or vascular access: given the objective of the ILCOR worksheets to look at cardiopulmonary resuscitation, the focus of the educational intervention should be the delivery of cardiopulmonary resuscitation to a patient/mannequin.

**Number of articles/sources meeting criteria for further review:**
37 articles were selected for further review (5 neonatal and 32 pediatric or adult).

Outcomes were organized as follows:

<table>
<thead>
<tr>
<th>A = Return of spontaneous circulation</th>
<th>C = Survival to hospital discharge</th>
<th>E = Other endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>B = Survival of event</td>
<td>D = Intact neurological survival</td>
<td></td>
</tr>
</tbody>
</table>

Additional outcomes were categorized according to Kirkpatrick’s levels of evaluation of educational programs (Donald Kirkpatrick's Learning Evaluation Model 1959):

- **E1 = Reaction.** These are subjective learner opinions of the educational intervention.
- **E2 = Learning.** These are objective assessments of the learner’s acquisition of knowledge, skills, or behaviours.
- **E3 = Transfer.** These are measures of practice change after completion of the learning experience.
- **E4 = Outcome.** These are measures of patient outcome as a consequence of learners receiving training.
## Summary of evidence

### Evidence Supporting Clinical Question

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair</td>
<td>Handley 2003 (E2)</td>
<td>Schrid 1999 (E2)§</td>
<td>Spooner 2007 (E2)</td>
</tr>
</tbody>
</table>

### Evidence Neutral to Clinical question

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Morgan 2002 (E2)</td>
<td>Iglesias-Vazquez 2007 (E2)</td>
<td>Knudson 2008 (E2)</td>
</tr>
<tr>
<td>Fair</td>
<td>Curran 2004 (E2)</td>
<td>Cherry 2007 (E2)</td>
<td>de Vries 2008 (E2)</td>
</tr>
</tbody>
</table>

Additional outcomes were categorized according to Kirkpatrick’s levels of evaluation of educational programs (Donald Kirkpatrick's Learning Evaluation Model 1959):

- **A** = Return of spontaneous circulation
- **B** = Survival of event
- **C** = Survival to hospital discharge
- **D** = Intact neurological survival
- **E** = Other endpoint
- **§** = Computer-based learning intervention or video self-instruction
- **Italics** = Neonatal studies

E1 = Reaction. These are subjective learner opinions of the educational intervention.

E2 = Learning. These are objective assessments of the learner’s acquisition of knowledge, skills, or behaviours.

E3 = Transfer. These are measures of practice change after completion of the learning experience.

E4 = Outcome. These are measures of patient outcome as a consequence of learners receiving training.
## Evidence Opposing Clinical Question

<table>
<thead>
<tr>
<th>Good</th>
<th>Fabius 1994 (E2)§</th>
<th>Isbye 2008 (E2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A = Return of spontaneous circulation  
C = Survival to hospital discharge  
E = Other endpoint  
B = Survival of event  
D = Intact neurological survival  
§ = Computer-based learning intervention or video self-instruction  
*Italics = Neonatal studies*

Additional outcomes were categorized according to Kirkpatrick’s levels of evaluation of educational programs (Donald Kirkpatrick’s Learning Evaluation Model 1959):  
E1 = Reaction. These are subjective learner opinions of the educational intervention.  
E2 = Learning. These are objective assessments of the learner’s acquisition of knowledge, skills, or behaviours.  
E3 = Transfer. These are measures of practice change after completion of the learning experience.  
E4 = Outcome. These are measures of patient outcome as a consequence of learners receiving training.
### REVIEWER’S FINAL COMMENTS AND ASSESSMENT OF BENEFIT / RISK:

37 controlled studies on resuscitation met inclusion criteria for this review.

The ideal study would ensure that the control and experimental groups spent equal time learning, and that the pretest and posttest methodology was relevant and masked to both learning interventions (see Appendix A); it would also measure both learning and clinical outcomes (levels E1 to E4 in the table). Few studies met this ideal.

A number of studies that did not include the core components of cardiopulmonary resuscitation (management of airway, breathing or circulatory collapse) were excluded. Studies were retained if a small portion of the learner performance included an aspect of cardiopulmonary resuscitation, but marked as “fair” or “poor”.

Studies that were simply pretest-posttest or descriptive designs were excluded from the review – as a rule they test whether more teaching results in more learning, and not whether the teaching through simulation is more effective than a control method. Even papers with apparent controls were excluded if the intervention provided significantly more training. For example, Wayne 2008 p56 LOE3(E3) gave the intervention group more training (and more immediate training, closer to the timing of outcome measurement) demonstrating improved adherence to protocols – this study was not included.

The nature of the research question also requires every included study to have a “control” and “intervention” group. This may be represented in observational studies by a regression analysis (Lee 2003 p651 LOE1(E2); Nackman 2003 p214 LOE 3(E2)).

The present state of cardiopulmonary resuscitation training is that the majority of courses have a component of “immersive learning” (whether “physical” simulation or “virtual” (computer or video)). As a consequence, most papers have “immersive learning” in both the control and experimental groups. This results in difficulty establishing which limb of each study is the “traditional” intervention, and which is “augmented” by simulation. For the purposes of this review, any “augmentation” of an established training process with feedback, debriefing, or computer-based learning is considered the experimental intervention.

Some studies tested two levels of technological fidelity, in which case the “lower” technological fidelity was taken as traditional training. Some studies looked at the effects of debriefing and feedback on learning; these interventions were taken as a “higher” technological fidelity as opposed to traditional.

Computer and video based learning are “virtual” simulations, and were therefore included in the review. These studies should be considered independently from “physical” simulations and have been marked (§) in the tables to clarify.

Most studies were randomized but the method of randomization was not always described.

Most studies looked at learner (E1) and course outcomes (E2) – as opposed to practice (E3) and patient (E4) outcomes. Most studies used overall scores in a simulated resuscitation as the outcome measure, whereas some used behavioural or technical measures (such as teamwork or measures of rate or depth of compression). Only 3 studies described a clinical outcome (Roberts 1997 p211 LOE2(E3); Knusden 2008 p263 LOE1(E3); Britt 2009 p533 LOE1(É, E4)) but are largely inadequately powered to demonstrate any significant difference.

Two cost efficiency studies point to the expense of simulated environments (de Vries 2008 p76 LOE1(E2); Iglesias-Vazquez 2007 p18 LOE 1(E2)).

Overall, given the inclusion and exclusion criteria, the studies were homogeneous in their design, but heterogeneous in their details: dividing participants into control or intervention, intervening, and evaluating outcome (using written test, evaluated simulation, and/or questionnaire).

Finally, an editorial note: the “fidelity” of an immersive learning experience can (and should) be divided into its “technological” fidelity (ie how close it is to physical reality) versus its “psychological” fidelity (ie how closely the learner perceives it to be to physical reality). It may be the perception of reality, or “psychological” fidelity that allows a learner to hone cognitive skills, whereas it may be the physical reality that trains technical skills – if this is the case, a review of simulation research should be sorted by the relative contribution of these two dimensions. No studies had the necessary details to perform this sorting.
Citation List


Reviewer's comments: Nursing students randomized to simulation vs “usual nursing courses” and assessed before and after by OSCE showed improved performance in the simulation group. Total recruited was 101 but 34 dropped out. The intervention was the simulation and the control appeared to be observation of that simulation. It was not clear what the control group did during that time.

Level of evidence: LOE 1
Relevance to question: Low – the nurses have limited exposure to cardiopulmonary resuscitation, and the scenarios largely test teamwork and collaboration.
Methodological quality: Poor (because of low relevance and high dropout)
Outcome(s) assessed: E2 – OSCE before and after the intervention
Magnitude of any observed effect: Approximately 6% improvement in OSCE score
Direction of support or otherwise for the question asked: Supportive


Reviewer's comments: A group of 36 postinternship physician graduates of ATLS courses received 45 minutes of additional airway simulation training based on the identified weaknesses of a historical control group (n=36). It is not clear whether that control group also received a similar amount of time on airway management. They were evaluated using simulated scenarios that required airway interventions. There appeared to be improved airway management in the intervention group.

Level of evidence: LOE 3 (not a true randomization of all participants)
Relevance to question: Fair – not sure if it tests the additional training or the simulation itself. Focuses on a task (airway management) that is essential to resuscitation training
Methodological quality: Poor, not sure that it is really testing simulation vs additional time spent learning.
Outcome(s) assessed: E2 – airway management in a simulation
Magnitude of any observed effect: Significant improvement in application of the task
Direction of support or otherwise for the question asked: Supportive – but limited in view of the uncertainty as to whether it is extra time spent on a task rather than the simulation per se that makes the difference


Reviewer’s comments: Junior residents in a trauma rotation were randomized to training in central line placement using standard (n=21) and simulated (n=13) methods. Multiple evaluations were done of observed “comfort” and “ability”, performance errors, and complications during real-life line placement. Although comfort and ability were apparently improved by simulation training, no statistically significant practice or patient outcomes were demonstrated.

Level of evidence: LOE 1 – the discrepant numbers of participants (21 vs 13) despite randomization are a concern.
Relevance to question: Relevant – looking to see whether simulation vs traditional teaching improves learner, practice, and patient outcomes.
Methodological quality: Poor – this study is inadequately powered to answer all these questions and it is uncertain what the primary outcome is.
Outcome(s) assessed: E2, E3 and E4 – There is an attempt to assess learner, practice and patient outcomes. The validity of assessment of learner outcome is uncertain
Magnitude of any observed effect: Probably insignificant, but small numbers do not allow conclusion
Direction of support or otherwise for the question asked: Neutral – this may be an issue with power and sample size

Reviewer's comments: First year residents (n=15) were randomized to learning NRP with a high technology manikin vs a standard manikin. There was no difference in written scores or intubation times. The intervention/simulation group rated the experience higher and required fewer redirects during the Megacode. This study may not be powered to demonstrate significant differences in some of these measures.

Level of evidence: LOE 1 – small number of participants makes it difficult to draw conclusions.

Relevance to question: Relevant – looking to see whether simulation vs traditional teaching improves learner outcomes.

Methodological quality: Poor – this study is inadequately powered to answer the questions and it is uncertain what the primary outcome is.

Outcome(s) assessed: E2, E1 – learner outcomes were assessed using the Megacode and written exam as well as learner rating of the experience.

Magnitude of any observed effect: Probably insignificant, but small numbers do not allow conclusion.

Direction of support or otherwise for the question asked: Supportive – if one takes the primary outcome as performance in the Megacode. Also supportive for subjective learner assessment.


Reviewer’s comments: Fifth year undergraduate medical students (n=45), following theoretical lecture, were randomized to learning neonatal resuscitation by self-study or by simulation. No difference was found in the written exams. I am not sure that the written exam is a good test of learning a psychomotor skill.

Level of evidence: LOE 1 – and all participants were accounted for.

Relevance to question: Borderline relevance – looking to see whether simulation vs self-directed learning improves learner outcomes using a written test.

Methodological quality: Poor – it is difficult to know whether a written test can evaluate learning that occurs in a simulation.

Outcome(s) assessed: E2 – learner outcomes were assessed using written test.

Magnitude of any observed effect: No effect observed.

Direction of support or otherwise for the question asked: Neutral – one would have preferred a psychomotor test.


Reviewer’s comments: First year residents (n=44) were randomized to simulation vs traditional training in an ATLS shock skills station, and evaluated by MCQ and OSCE. No significant difference was found in performance. It is not clear from the paper exactly which skills were evaluated in the shock scenarios.

Level of evidence: LOE 1

Relevance to question: Uncertain, as it is unclear which skills were being tested.

Methodological quality: Fair – the question is clearly stated and the design should provide the answer.

Outcome(s) assessed: E2 – learner outcomes were assessed using written test and OSCE.

Magnitude of any observed effect: No effect observed.

Direction of support or otherwise for the question asked: Neutral – it is assumed that management of shock means this study meets inclusion criteria require training in some aspect of resuscitation. Clarification of the training and OSCE content would help.


Reviewer's comments: Final year medical students non-randomized (convenience) to be taught by traditional ACLS (n=38) vs traditional + multimedia ACLS (n=75) were evaluated and found to perform similarly on knowledge. The MM group had greater difficulty on technical skills. This is a simulation vs simulation plus computer-based learning (CBL) question.

Level of evidence: LOE 2

Relevance to question: Low, as the addition of a multimedia component to ACLS may not enhance the simulation.

Reviewer's comments: Medical students (note high dropout 29 out of 60) randomized to a 4-month “booster” with simulation vs video and then assessed at 8 months by simulation showed no differences in knowledge, skills or confidence.

Level of evidence: LOE 1
Relevance to question: High
Methodological quality: Fair – high dropout rate resulted in inadequate power
Outcome(s) assessed: E2 – retention of course content tested using resuscitation simulation
Magnitude of any observed effect: Minimal, but may be inadequately powered to demonstrate an effect
Direction of support or otherwise for the question asked: Neutral


Reviewer's comments: Comparing traditional AED training with self-directed learning for randomized nurses showed no difference in overall performance – some statistically significant but clinically insignificant differences were seen. Fifteen nurses in each group.

Level of evidence: LOE 1
Relevance to question: High
Methodological quality: Fair – may have been inadequately powered to show a difference. Assessors were not masked to the intervention.
Outcome(s) assessed: E2 – performance in a simulated scenario
Magnitude of any observed effect: Minimal, apart from the significant cost-saving of self-directed learning
Direction of support or otherwise for the question asked: Neutral


Reviewer’s comments: Nurses randomized to debriefing with or without audiovisual feedback on a simulator (note dropout rate was 15 out of 65) were evaluated using a simulator for adult CPR. Debriefing and feedback improve performance of CPR and the two effects are additive. Not sure this is really testing simulation versus traditional unless you consider debriefing and audiovisual feedback as unique to simulation rather than traditional teaching methods.

Level of evidence: LOE 1
Relevance to question: Low – depends on whether the addition of debriefing and audiovisual feedback can be considered an augmented simulation.
Methodological quality: Fair, high dropout, otherwise interesting design.
Outcome(s) assessed: E2 – performance during a simulated resuscitation
Magnitude of any observed effect: Significant
Direction of support or otherwise for the question asked: Supportive – positive for increased feedback and debriefing, evaluating chest compressions rate and depth

Pediatric residents were randomized to high technical fidelity simulation (n=25) vs low technical fidelity training (n=24) during a PALS training course, and evaluated using cognitive scoring during standard scenarios. Participants trained using simulation showed a greater posttest increase in course scores.

**Reviewer’s comments**

**(Pediatric residents were randomized to high technical fidelity simulation (n=25) vs low technical fidelity training (n=24) during a PALS training course, and evaluated using cognitive scoring during standard scenarios. Participants trained using simulation showed a greater posttest increase in course scores.**

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>LOE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance to question</td>
<td>Relevant</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>Good – the question is clearly stated and the design should provide the answer. The validity of the assessment tool is unclear</td>
</tr>
<tr>
<td>Outcome(s) assessed</td>
<td>E2 – learner outcomes were assessed using simulated scenarios</td>
</tr>
<tr>
<td>Magnitude of any observed effect</td>
<td>Significantly greater improvement in scores in the simulated group compared to the traditional</td>
</tr>
<tr>
<td>Direction of support or otherwise for the question asked</td>
<td>Supportive</td>
</tr>
</tbody>
</table>


**Reviewer’s comments**

70 subjects, randomized to “traditional” or “computer” education and re-evaluated 6 months later by instructor. Results favoured the traditional method. 16 out of 70 subjects lost to follow-up. This is a CBL intervention.

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>LOE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance to question</td>
<td>High – but again, the traditional and computer interventions are both simulations, so there is difficulty determining which is the experiment and which is the control.</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>Good – but note high dropout rate</td>
</tr>
<tr>
<td>Outcome(s) assessed</td>
<td>E2 – Both learner satisfaction and simulated resuscitation after 6 months.</td>
</tr>
<tr>
<td>Magnitude of any observed effect</td>
<td>No difference on most parameters, but higher fail rate on CBL</td>
</tr>
<tr>
<td>Direction of support or otherwise for the question asked</td>
<td>Opposing – this assumes that the CBL was the intervention.</td>
</tr>
</tbody>
</table>


**Reviewer’s comments**

First year pediatric interns (n=38) were randomized to training specific tasks including bag-mask ventilation (BMV) using simulation vs traditional (experiential) learning. Skills testing by simulation occurred immediately after intervention and 7 months later. No differences were found between groups for BMV (although there were some differences in non-resuscitation skills).

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>LOE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance to question</td>
<td>Relevant, but not primarily related to resuscitation</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>Fair – it is difficult to know whether the intervention group had more training. There were a number of outcomes with no clear definition of a primary outcome</td>
</tr>
<tr>
<td>Outcome(s) assessed</td>
<td>E2 – learner outcomes were assessed using simulated skills stations</td>
</tr>
<tr>
<td>Magnitude of any observed effect</td>
<td>No observed effect on resuscitation skills. There was some effect on IV insertion and lumbar puncture</td>
</tr>
<tr>
<td>Direction of support or otherwise for the question asked</td>
<td>Neutral</td>
</tr>
</tbody>
</table>


**Reviewer’s comments**

36 general nurses randomized to training on a simulator with or without auditory feedback. Experimental group performed better at inflations and chest compressions

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>LOE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance to question</td>
<td>Low – this is testing a simulator against and augmented simulator</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>Fair – not masked</td>
</tr>
<tr>
<td>Outcome(s) assessed</td>
<td>E2 – testing effectiveness of ventilation and chest compressions in the simulator</td>
</tr>
<tr>
<td>Magnitude of any observed effect</td>
<td>Substantial</td>
</tr>
</tbody>
</table>
Direction of support or otherwise for the question asked: Supportive


<table>
<thead>
<tr>
<th>Reviewer's comments</th>
<th>Supportive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multifaceted health care providers (n=53) were randomized to low or high technical fidelity ACLS training. There was no significant improvement in posttest scores or skills, although participants subjectively rated the high fidelity as recommendable.</td>
<td></td>
</tr>
<tr>
<td>Level of evidence</td>
<td>LOE 1</td>
</tr>
<tr>
<td>Relevance to question</td>
<td>Relevant</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>Good – it is difficult to know whether the intervention group had more training. There were a number of outcomes with no clear definition of a primary outcome</td>
</tr>
<tr>
<td>Outcome(s) assessed</td>
<td>E2 – learner outcomes were assessed using simulated skills stations</td>
</tr>
<tr>
<td>Magnitude of any observed effect</td>
<td>No significant effect on resuscitation knowledge or skills</td>
</tr>
<tr>
<td>Direction of support or otherwise for the question asked</td>
<td>Neutral</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Reviewer's comments</th>
<th>Comparing standard fidelity ALS with high technological fidelity ALS for randomized health professionals (some police and fire officers) to look at cost efficiency and pass rate in the ALS exam. The abstract is incorrect as it states the p-value is &lt;0.01 when the pass rate is compared. There was an insignificant improvement in pass rate (78% vs 88%, p = 0.06) implying no difference or a trend.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of evidence</td>
<td>LOE 1</td>
</tr>
<tr>
<td>Relevance to question</td>
<td>High</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>Good – although some participants were not healthcare professionals</td>
</tr>
<tr>
<td>Outcome(s) assessed</td>
<td>E2 – pass rate for the course and cost per pass</td>
</tr>
<tr>
<td>Magnitude of any observed effect</td>
<td>No significant change in pass rate, but marked difference in cost (392 vs 1320 Euros)</td>
</tr>
<tr>
<td>Direction of support or otherwise for the question asked</td>
<td>Neutral</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Reviewer's comments</th>
<th>Second year medical students randomized to training ALS with an instructor vs with automated audio feedback on a simulator were evaluated using a 2 min simulation for adult CPR at the end of the course and 3 months later. No difference was seen between groups except in the performance of bag-mask ventilation which was better following human instruction.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of evidence</td>
<td>LOE 1</td>
</tr>
<tr>
<td>Relevance to question</td>
<td>Low – it is uncertain which should be the intervention, the simulation with the instructor or the simulation with the voice advisory mannequin (VAM)</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>Fair</td>
</tr>
<tr>
<td>Outcome(s) assessed</td>
<td>E2 – CPR performance in a simulation</td>
</tr>
<tr>
<td>Magnitude of any observed effect</td>
<td>Significant</td>
</tr>
<tr>
<td>Direction of support or otherwise for the question asked</td>
<td>Opposing – traditional training with an instructor was superior</td>
</tr>
</tbody>
</table>

Reviewer’s comments: Final family practice residents randomized to a 4-month booster with hands-on, video or control (video, n = 13; hands-on, n = 14; control, n = 17) were evaluated 3 to 5 months later and found to perform similarly on knowledge and practice (simulated). Not sure that the power was there to show a difference.

Level of evidence: LOE 1
Relevance to question: High – asking whether simulated experiences can boost retention
Methodological quality: Poor, inadequately powered.
Outcome(s) assessed: E2 – both written exam and simulated resuscitation.
Magnitude of any observed effect: Low
Direction of support or otherwise for the question asked: Neutral


Reviewer’s comments: Surgical residents randomized to training trauma resuscitation (not much medical resuscitation) by didactic, scenario-based lecture vs simulation were evaluated using written test and real-life videorecording at the end of the course (note not all data included). They developed their own evaluation tools and attempted to validate using inter-rater reliability assessments. No significant difference was seen between groups except in teamwork (p<0.04), which was better following simulation – simulation scores tended to be higher, suggesting a benefit in crisis management outcomes.

Level of evidence: LOE 1
Relevance to question: High – but do we use didactic teaching as our controls?
Methodological quality: Good for course outcomes, fair for teamwork in real-life videorecording
Outcome(s) assessed: E2 – written tests; E3 – videorecording of real-life performance
Magnitude of any observed effect: Teamwork effect was noted in a sub-group analysis
Direction of support or otherwise for the question asked: Neutral for practice outcomes, supportive for teamwork (essentially a sub-group of resource management)


Reviewer’s comments: Comparing traditional methods with high fidelity for historically controlled senior internal medicine residents in PGY1 and tested in PGY3 showed improved performance in a simulation.

Level of evidence: LOE 3
Relevance to question: Difficult to know whether the intervention was “additional” training, as opposed to simulation being an alternative to traditional training.
Methodological quality: Poor, no correction for confounders and 2 years between groups. Control may have received less training time.
Outcome(s) assessed: E2 – performance in a simulated resuscitation
Magnitude of any observed effect: Significant
Direction of support or otherwise for the question asked: Supportive


Reviewer’s comments: Surgery interns (n=60) randomized to simulation training vs moulage in trauma assessment (adult) and then randomly assessed using simulation or moulage (manikin and actors) by face-to-face or video evaluators were found to perform better after simulation. Very neat design, but resuscitation only a small part of the assessment.

Level of evidence: LOE 1
Relevance to question: Difficult to assign, as simulated patients are also “simulation”
Methodological quality: Good, not sure that it answers the question of simulation vs traditional as “moulage” may not be considered “traditional”
Outcome(s) assessed: E2 – observed assessment objectives in a simulated scenario
Magnitude of any observed effect | Significant statistically, but of questionable clinical significance.
---|---
Direction of support or otherwise for the question asked | Supportive, in favour of simulation without actors.


Reviewer’s comments | Multiprofessional ACLS course participants randomized to training using manikins vs manikins plus live actors were evaluated using written test (MCQ) at the end of the course and 6 months later (note dropout). No significant difference was seen between groups.
---|---
Level of evidence | LOE 1
Relevance to question | Difficult to ascertain as the comparison is between simulation and augmented simulation
Methodological quality | Poor, data only available for 225 out of 435 participants.
Outcome(s) assessed | E2 – knowledge only.
Magnitude of any observed effect for the question asked | None
Direction of support or otherwise for the question asked | Neutral


Reviewer’s comments | Medical students (n=144) randomized to facilitated video learning vs simulation were found to perform equally well on simulation and written tests. Not sure that this study should be included as it looks at hands-on simulation vs video simulation – but there are a number of studies like this that contribute to the question, by using video as a control intervention compared to simulation.
---|---
Level of evidence | LOE 1
Relevance to question | Fair, as this is comparing simulation with video (is that traditional?)
Methodological quality | Good, not sure that it asks the question of traditional vs simulation as it compares two forms of simulation (visual and experiential).
Outcome(s) assessed | E2 – a written test evaluates knowledge. E1 – subjective measure of experiential versus visual learning
Magnitude of any observed effect for the question asked | Of questionable clinical benefit.
Direction of support or otherwise for the question asked | Neutral for course measures. Supportive for subjective learner evaluation of simulation


Reviewer’s comments | Surgery interns (n=86) divided in a non-randomized fashion to simulation training vs small group sessions in trauma care (adult) and then assessed by OSCE. Multiple regression analysis to adjust for confounders suggests that training on the simulator predicts a higher score in the OSCE.
---|---
Level of evidence | LOE 3
Relevance to question | High relevance
Methodological quality | Fair, complex design and confounders corrected by regression analysis.
Outcome(s) assessed | E2 – performance in an OSCE
Magnitude of any observed effect for the question asked | Statistically significant
Direction of support or otherwise for the question asked | Supportive


Reviewer’s comments | Pediatric residents (28) with historical controls (30) taught pediatric resuscitation with an augmented course with feedback vs a traditional course assessed using a “fund of knowledge”
were found to perform better after augmentation: that extra training results in better outcomes is not a surprise. It is likely that both limbs of the study had simulation components.


**Reviewer’s comments**
Anesthesia trainees (n=40) quasi-randomized to computer-screen simulation vs high fidelity simulator were found to perform equally well on simulation and testing. Difficult to ascertain randomization to groups – I suspect not at all to computer or manikin, but to scenario. Not a great design – “resuscitation” skills were limited in the scenarios, which were mostly about anesthesia management. This is a CBL vs “simulation” comparison.


**Reviewer’s comments**
CSBT, CSBT plus simple manikins, and CSBT plus high fidelity manikins were compared using simulated scenarios to evaluate, randomizing 64 (20 did not attend, and 3 dropped out) “medical officers”. There was some improvement in the group trained with high fidelity. This study uses CBL as its baseline and adds increasing fidelity of simulation.


**Reviewer’s comments**
Randomized controlled study – method of randomization unclear, and no power analysis. Not blinded. All participants accounted for, but unclear how there were 32 in the control (manikin only training) group and 20 in the manikin and OR training group. Simulation and real-life training in LMA placement by nurses in adults had equivalent results when tested in real-life. This suggests that this specific skill may be taught using simulation.

Reviewer’s comments: The study was a nonrandomized, nonequivalent comparison group design of a convenience sample of 86 ACLS students using interactive videodisc vs traditional teaching (ACLS instructor face-to-face with simulation) to learn airway management. No difference was found in performance – there was a subgroup of first-time ACLS participants who performed better in one task after IVD, suggesting that these learners may have benefited. This study appears to be a (psychomotor) simulation vs CBL (virtual) simulation

Level of evidence: LOE2 – non-randomized control group
Relevance to question: Low relevance – a CBL vs simulation study
Methodological quality: Poor – convenience sample without power analysis
Outcome(s) assessed: E2 – performance in a mock scenario
Magnitude of any observed effect: No effect on primary outcome. Subgroup analysis suggests that CBL may improve performance in first-time ACLS participants
Direction of support or otherwise for the question asked: Neutral


Reviewer’s comments: Anesthesia residents randomized to learn using a computerized ACLS simulation program (n=23) vs textbooks (n=22) in preparation for retesting in a standardized Mega Code were found to perform better after computer simulation. More individuals studied the simulation but results the same if those who did not study excluded. This is computer-based learning (CBL) vs traditional.

Level of evidence: LOE 1
Relevance to question: Limited – compares CBL with text
Methodological quality: Fair, some did not complete the control intervention, which may have affected outcome
Outcome(s) assessed: E2 – performance in a simulated code
Magnitude of any observed effect: Significant from and intention-to-treat perspective
Direction of support or otherwise for the question asked: Supportive


Reviewer’s comments: Anesthesia residents randomized to computer assisted learning (n=16) vs text (n=15) assessed using simulated scenarios were found to perform better after computer training. This is CBL vs text.

Level of evidence: LOE 1
Relevance to question: Limited, compares CBL to text
Methodological quality: Good
Outcome(s) assessed: E2 – performance in a simulated code
Magnitude of any observed effect: Statistically significant
Direction of support or otherwise for the question asked: Supportive


Reviewer’s comments: Multiprofessional ER teams (n=4) were randomized to 8 hours of training in teamwork and simulation vs a standard ER shift. External observers rated team behaviours in the ER with a validated tool. No differences were seen in the two groups.

Level of evidence: LOE 2 – given the low number of teams, despite randomization

Reviewer's comments: Comparing traditional manikin with high fidelity (provides feedback to operator) for randomized first year healthcare students improved performance in the procedure and 6 weeks later (dropout from 98 to 66 students). This study tests a “task trainer”.

Level of evidence: LOE 1
Relevance to question: Relevant
Methodological quality: Fair, high dropout rate. Outcome of chest compressions.
Outcome(s) assessed: E2 – the ability of trainees to provide chest compression rate and depth
Magnitude of any observed effect: Significant
Direction of support or otherwise for the question asked: Supportive


Reviewer's comments: Medical students randomized to facilitated learning (heart failure or anaphylaxis) using screen-based simulation (n=33) vs facilitated lectures (n=31) were evaluated using a human patient simulator several days later. No significant difference was seen between groups in all but one area. This study compares CBL with traditional teaching. Many of the interventions were not cardiopulmonary resuscitation

Level of evidence: LOE 1
Relevance to question: Relevant for CBL as a simulation
Methodological quality: Fair
Outcome(s) assessed: E2 – evaluation on a simulated scenario
Magnitude of any observed effect: Minimal
Direction of support or otherwise for the question asked: Neutral


Reviewer's comments: Comparing traditional NRP training with NRP plus team training for randomized interns improved team behaviours during a simulated evaluation. Of 40 randomized, 32 were studied. Not sure that this tells us anything as more team training results in more team behaviours.

Level of evidence: LOE 1
Relevance to question: Limited
Methodological quality: Poor, shows that more team training results in more team behaviours.
Outcome(s) assessed: E2 – team behaviours
Magnitude of any observed effect: Significant
Direction of support or otherwise for the question asked: Supportive for team behaviours.


Reviewer's comments: Incoming freshmen medical students randomized to be taught by video self-instruction (VSI) vs
traditional AHA Heartsaver course were evaluated using CPR simulation and found to perform the same or better after VSI. It is difficult to be sure which is the simulation and which is traditional – it appears that the VSI with mannequin is the intervention.

Level of evidence
Relevance to question
Methodological quality
Outcome(s) assessed
Magnitude of any observed effect
Direction of support or otherwise for the question asked

LOE 1
Relevant – the VSI and mannequin constitutes a self-instructed simulation.
Good
E2 – performance in a CPR simulation
Significant improvement in pass rate for VSI
Supportive


Reviewer's comments
Internal medicine second-year residents (n=38) randomized to baseline then simulation vs simulation then baseline (crossover design). Solid design.

Level of evidence
Relevance to question
Methodological quality
Outcome(s) assessed
Magnitude of any observed effect
Direction of support or otherwise for the question asked

LOE 1
High
Good
E2 – Performance in a simulated code
Significant (p<0.05).
Supportive


Reviewer's comments
Internal medicine second- (n=38) and third-year residents (n=40) were compared, the latter being historical controls. The second-year residents received additional 10 hours of simulation-based training – whereas more time had passed since the third-year residents had traditional training. Resuscitation flow charts were abstracted for key interventions that demonstrated adherence to ACLS protocols, and teams led by the two different cohorts were compared. The results showed more adherence to protocols by the group who had the more recent additional training. This supports the premise that more training and more recent training improves practice outcomes.

Level of evidence

LOE 3
Relevance to question: High
Methodological quality: Poor – there was a convenience sample with no randomization and use of historical controls with a different level of training. The intervention group actually had more training, and it was more recent. The analysis was by logistic regression controlling for patient factors.
Outcome(s) assessed: E3 – practice was evaluated by data abstraction from patient resuscitation flow charts
Magnitude of any observed effect: Residents who participated in simulation training were significantly more likely to adhere to ACLS protocols
Direction of support or otherwise for the question asked: Excluded – this study suggests that more training and more recent training will result in better adherence to ACLS protocols. Simulation is not necessarily the reason for this improvement.


Reviewer's comments: Randomized crossover design. Paramedic students (n=24) randomized to CPR simulator training with voice feedback vs traditional CPR simulator training assessed using the simulator were found to perform better with voice feedback. This is testing a feedback feature of the simulator – “task trainer”.
Level of evidence: LOE 2 (not a true randomization of all participants)
Relevance to question: Low, testing task trainer
Methodological quality: Poor, not sure that it is really testing simulation vs traditional or a device.
Outcome(s) assessed: E2 – Performance of the task.
Magnitude of any observed effect: Significant
Direction of support or otherwise for the question asked: Supportive, technical endpoint of amount and rates of compressions and ventilation

Appendix A:
**IDEAL STUDY DESIGN conditions should include:**
- All Kirkpatrick's levels of evaluation should be satisfied.
- Pretest-posttest evaluators should be masked to the randomization.
- Method of evaluation should be familiar to both the control and the intervention.
- Control and intervention groups should have the same hours of intervention.
- Delayed posttest should evaluate decay of learning/practice

---

Ideal study design for educational program evaluation (Aziz 2009) v1.0