Supplemental Digital Content 2

Explanation and Elaboration of the Simulation-Specific Extensions for the CONSORT and STROBE Statements

In this document, we provide examples for each of the items where a simulation-specific extension was created for the CONSORT and STROBE Statements. Examples were chosen to reflect ideal reporting for the item and the associated simulation-specific extension. After each example, an explanation is provided that focuses on providing rationale and further elaboration for the simulation extension. We refer the reader back to the CONSORT[1] and STROBE[2] explanation and elaboration documents for further details related to the original items.

Item: Title and Abstract
CONSORT: 1a: Identification as a randomized trial in the title; 1b: Structured summary of trial design, methods, results, and conclusions
STROBE: (a) Indicate the study's design with a commonly used term in the title or the abstract; (b) Provide in the abstract an informative and balanced summary of what was done and what was found.

Extension: In abstract or key terms the MESH or searchable keyword term must have the word "simulation" or "simulated".

Examples
CONSORT:
"Abstract: PURPOSE: To compare pelvic ultrasound simulators (PSs) with live models (LMs) for training in transvaginal sonography (TVS). METHOD: The authors conducted a prospective, randomized controlled trial of 145 eligible medical students trained in TVS in 2011-2012 with either a PS or an LM. A patient educator was used for LM training. Simulated intrauterine and ectopic pregnancy models were used for PS training. Students were tested using a standardized patient who evaluated their professionalism. A proctor, blinded to training type, scored their scanning technique. Digital images were saved for blinded review. Students rated their training using a Likert scale (0 = not very well; 10 = very well). The primary outcome measure was students' overall performance on a 40-point assessment tool for professionalism, scanning technique, and image acquisition. Poisson regression and Student t test were used for comparisons. RESULTS: A total of 134 students participated (62 trained using a PS; 72 using an LM). Mean overall test scores were 56% for the PS group and 69% for the LM group (P = .001). A significant difference was identified in scanning technique (PS, 60% versus LM, 73%; P = .001) and image acquisition (PS, 37% versus LM, 59%; P = .001). None was observed for professionalism. The PS group rated their training experience at 4.4, whereas the LM group rated theirs at 6.2 (P < .001). CONCLUSIONS: Simulators do not perform as well as LMs for training novices in TVS, but they may be useful as an adjunct to LM training.

MESH terms: Clinical Competence, Education, Medical, Undergraduate/methods, Female, Humans, Models, Anatomic, Patient Simulation, Ultrasonography, Uterus/ultrasonography."[3]

STROBE:
"Abstract: INTRODUCTION: Medical school graduates are expected to possess a broad array of clinical skills. However, concerns have been raised regarding the preparation of medical students to enter graduate medical education. We designed a simulation-based "boot camp" experience for students entering internal medicine residency and compared medical student performance with the performance of historical controls who did not complete boot camp. METHODS: This was a cohort study of a simulation-based boot camp educational intervention. Twenty medical students completed 2 days (16 hours) of small group simulation-based education and individualized feedback and skills assessment. Skills included (a) physical examination techniques (cardiac auscultation); technical procedures including (b) paracentesis and (c) lumbar puncture; (d) recognition and management of patients with life-threatening conditions (intensive care unit clinical skills/mechanical ventilation); and (e) communication with patients and families (code status discussion). Student posttest scores were compared with baseline scores of postgraduate year 1 (PGY-1) historical controls to assess the effectiveness of the intervention. RESULTS: Boot camp-trained

medical students performed significantly better than PGY-1 historical controls on each simulated skill (P<0.01). Results remained significant after controlling for age, sex, and US Medical Licensing Examination step 1 and 2 scores (P<0.001). CONCLUSIONS: A 2-day simulation-based boot camp for graduating medical students boosted a variety of clinical skills to levels significantly higher than PGY-1 historical controls. Simulation-based education shows promise to help ensure that medical school graduates are prepared to begin postgraduate training.

MESH terms: Adult, Clinical Competence/standards, Cohort Studies, Curriculum, Education, Medical, Graduate, Female, Humans, Internal Medicine/education, Male, Patient Simulation, Students, Medical, Teaching/methods, United States, Young Adult."[4]

Explanation

Simulation-based research appears in both simulation-specific journals and healthcare research journals across a variety of domains. The use of simulation as a modality for, or subject of research, should be clarified in the abstract (and/or title) to enable efficient and accurate searches in journal databases. As journals have different formats for abstracts, not all abstracts will have intuitive headings that allow a reader to quickly skim through and identify the study design or its use of simulation. Adding the word "simulation" or "simulated" or "simulator" within the abstract and/or title allows researchers, clinicians, and educators to quickly access the manuscript based on its modality, similar to searching for study methodology - whether randomized control trial or observational. It is important to recognize that **"**simulation", "simulated" and "simulator" are not currently MeSH terms but "Patient Simulation" and "Computer Simulation" are recognized MeSH terms. Only "Patient Simulation" (I02.903.525) is related to the MeSH taxonomy of education and teaching[5]. This MeSH term was introduced in 1992. "Computer Simulation" (L01.224.160) is related to the Information Science tree (L01).

Item: Introduction/Background

CONSORT: 2a: Scientific background and explanation of rationale; 2b: Specific objectives or hypotheses
STROBE: Explain the scientific background and rationale for the investigation being reported.

Extension: Clarify whether simulation is subject of research or investigational method for research.

Examples

CONSORT:

"Introduction/background: ...the scientific study of Emergency Department Procedural Sedation (EDPS) safety is hampered by a lack of consistent definitions, the relatively low reported incidence of adverse events, and the variety of clinical practice models and pharmacologic agents... EDPS can be conceptually modeled in a manner that fits well with simulation-based investigation. This research program explored the following objectives: (1) to assess EDPS provider performance through in situ simulation and (2) to concurrently develop and study the effect of an experimental just-in-time safety system."[6]

STROBE:

"Introduction/background: Distributing learning activities over time (distributed practice) is a key instructional design strategy promoting skill improvement that was not included in our previous trials … This education strategy can be subdivided into experiences that are not only apportioned across time but also offered **"**just in time,**"** or immediately before the clinical task or procedure being trained for, and **"**just in place,**"** when the learning experience occurs in the actual workplace. The described intervention occurred both just in place and just in time (JIPT). Our hypothesis was that the addition of JIPT simulation immediately before infant LP attempts would have a greater impact on interns**'** clinical LP success rate than a solitary training session."[7]

Explanation

Simulation-based research has two main categories: 1) simulation as an educational intervention within healthcare (subject of research), and 2) simulation as investigational methodology[8]. These exact words need not be used in all papers, however it should be clarified in the introduction and/or background which of these two categories of simulation research was conducted. Both the CONSORT and STROBE Statements require an appropriate scientific background and rationale for the study, to demonstrate a gap in the science. The first example above provides an example of a randomized clinical trial utilizing simulation

as an investigational method, and the authors provided a background on the scientific knowledge to date on the chosen healthcare topic. The second example describes the study of simulation as the subject of research. This study describes an instructional design component (distributed practice) that will be examined in the study. The background can also concentrate on the available simulation-based and non-simulation based interventions that are standard of care and/or are considered best educational practice. Authors should carefully clarify the weaknesses of the standard interventions and why the intervention under study is likely to improve outcomes. This may come in the form of evidence-based studies or conceptual frameworks.

Item: Intervention (CONSORT) and Variables (STROBE)
CONSORT: The interventions for each group with sufficient details to allow for replication, including how and when they were actually administered.
Extension: Describe the theoretical and/or conceptual rationale for the design of the intervention. Clearly describe all simulation-specific exposures, potential confounders, and effect modifiers.

STROBE: Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.
Extension: Describe the theoretical and/or conceptual rationale for the design of the intervention / exposure. Describe the intervention / exposure with sufficient detail to permit replication. Clearly describe all simulation-specific exposures, potential confounders, and effect modifiers.

Examples
CONSORT:
"Methods:
Scripted vs. Non-Scripted Debriefing: All novice instructors received the scenario 2 weeks prior to the study session. Instructors randomized to scripted debriefing were also given the script with no instruction on how to use it except with direction to use and follow the script as closely as possible during the debriefing. Instructors randomized to non-scripted debriefing were asked to conduct a debriefing to cover the pre-defined learning objectives, with no specific instruction on what style or method of debriefing to use. All instructors held a clipboard while observing the simulation session; to hold the debriefing script and take notes. This allowed for blinding of the video reviewers as to non-scripted vs scripted debriefing. A research assistant verbally intervened to stop the debriefings that extended to 20 minutes.
High vs. Low Physical Realism Simulators: A pre-programmed Laerdal Simbaby™ infant simulator was used for all simulation sessions. To create "high" physical realism (HiR), full simulator functions were activated ("turned on") including vital sign monitoring, audio feedback, breath sounds, chest rise, heart sounds and palpable pulses. "Low" physical realism (LoR) groups had the identical simulator but the compressor was "turned off", thus eliminating physical findings described above. In addition, the LoR simulator was connected to a monitor, but it only displayed the cardiac rhythm, and not pulse oximetry, respiratory rate, blood pressure, temperature and audio feedback present in the HiR group. All other aspects of the simulated resuscitation environment were standardized. See eMethods for details.
Simulation Scenario: The 12 minute scenario was divided up into 3 separate stages (hypotensive shock, ventricular fibrillation, return to normal sinus rhythm), with progression from one stage to the next at pre-determined time intervals (2 minute for first transition, 10 minutes for second transition) irrespective of how the team managed the patient (eTable 2). The scenario was stopped at a maximum of 12 minutes, or earlier if the medical team felt a palpable pulse and verbalized a normal sinus rhythm. Specific cues were delivered by the research assistant for the low realism scenarios when specifically asked for by team members per a standardized script (eg. Level of consciousness, saturations, blood pressure, heart sounds etc). Only a few cues were delivered by research assistant for high realism scenarios (eg. mottled appearance, capillary refill), specifically in instances when the simulator was unable to provide realistic feedback....
Simulated Environment: Simulation scenarios were conducted in the simulation rooms at the various recruitment sites, with rooms closely mimicking the clinical work environment. A detailed equipment list was provided to all recruitment sites to ensure resource availability was standardized across all sites. Medication availability was also standardized and limited to 8 different resuscitation drugs. A standardized intravenous setup was used at all recruitment sites which allowed injection of fluid / medication during the scenario. Placement of the code cart relative to the stretcher was standardized (at the foot of the bed, within

3 feet of the stretcher). Subjects were permitted to carry and use PALS pocket cards / cognitive aids during the scenarios. Additional medical test such as electrocardiograms, radiographs and bloodwork could be ordered by team members but results were not made available for review.
Debriefing Script: A debriefing script was designed for novice instructors to facilitate a 20-minute debriefing session … The language used in the script was developed based on the debriefing theory known as "advocacy-inquiry" … the content of the script was divided up into 2 main topics: medical management and crisis resource management (CRM) ... The script provides specific phrases, in the model of advocacy-inquiry, for each key intervention or task, including options for if the task was "performed well" or "needed work". The script then guides the facilitator through the debriefing process by suggesting follow-up phrases or questions … The debriefing script was included as eTable 3, 4,5 of this paper."[9]

STROBE:
"Methods:
We conducted a cross sectional survey and a prospective observational cohort study of sCPAs. Simulated Cardiopulmonary Arrests (sCPAs): Upon arrival at the simulation lab, each resident received a standardized orientation to the human patient mannequin simulator, the Laerdal SimMan®. They were told that when they re-entered the simulated hospital room, there would be two individuals acting as their nurses who would be helpful but would not share independent ideas on how to manage the "patient". Residents were instructed to ask for information, tests, personnel or equipment they would normally want in order to manage their patient, and the team would simulate having more personnel if necessary. The standard resuscitation equipment used throughout our hospital was readily available within the room, including a Zoll M series® semi-automatic defibrillator. Upon the resident's re-entry into the room, one of the nurses stated, "We asked you to come see this patient because he is having PVCs (premature ventricular contractions). He is a 12 year old who came up from the Emergency Department (ED) approximately 1 hour ago. He came to the ED because he was a little short of breath. His labs were hemolyzed except a creatinine of 2.0. He has been having a few PVCs per minute." SimMan® was programmed with identical vital signs for every pediatric resident's mock code. The mannequin would answer questions if asked. At 1 minute, the patient became unresponsive, apneic and pulseless due to onset of PVT. The script included a standardized progression depending upon the resident's actions. If the resident defibrillated the patient four times, (i.e. shock, shock, shock, epinephrine, shock – per 2000 AHA guidelines) the fourth shock converted the cardiac rhythm to sinus bradycardia, but with no pulse, i.e., "Pulseless Electrical Activity" (PEA). If the resident delivered epinephrine after development of PEA, the patient regained a palpable pulse and measurable blood pressure … The sCPA ended when either: 1) the patient regained sinus rhythm with a pulse or 2) the sCPA had run for 15 minutes after onset of PVT. While physiologically, the patient would likely require compressions in order to be resuscitated, we did not require compressions for the scenario to progress in order to assess the resident's recall of the correct procedures."[10]

<u>Explanation</u>
Simulation-based research affords investigators unique opportunities to control or measure many elements of the study design whether it includes components of an intervention or variables in a non-interventional study (provider, patient and systems). Investigators should report the elements they have controlled, as well as specific methods of how they controlled these elements and describe elements that they did not control. The theoretical and/or conceptual rationale for the design and approach to controlling these elements should be described in the methods section. Additionally the outcomes, exposures, confounders should be clearly described in terms of both how and why they were controlled or not controlled. Independent of the study design it is important that the studies exploring simulation as the intervention clearly describe all elements of the simulation-based intervention. This should include both the area of inquiry or comparison and explicit description of other factors that could impact outcomes. For example, a study comparing the efficacy of two different simulator types must describe all of the other factors including participant orientation, the simulation event, scenario(s), environment and instructional design approaches (Table 3). The descriptions of these elements should be provided in sufficient detail that the study can be replicated by other investigators. In some cases, this may require online appendices that serve as a supplement to the manuscript. For example, a study could suggest a specific approach to debriefing and require training for facilitators or could require the use of scripted debriefings with video reviews to ensure adherence to the study protocol. This level of detail will often require an appendix including a description of the scenario, debriefing script and/or technical specifications of the simulator that may

exceed the word limit of the manuscript. Methods and strategies used to control all other simulation specific variables (Table 3) that are not part of the intervention should be reported in the methods.

Research using simulation as an investigative modality providers researchers the ability to answer questions that may not be feasible or ethical to address with other research methodologies[8]. Simulating patients and/or providers has the potential to provide researchers control over nearly every aspect of the simulated environment to mirror specific research targets in the healthcare system. The degree of control implemented for each element should be clearly described with detailed descriptions of all elements that were controlled. The variables that are not controlled in the study should also be reported. For example, studies examining human factors at the individual or team level would require detailed descriptions of the patient and work environments. A study examining the efficacy of a novel medical device should include clear descriptions of how the participants where oriented to use the device, the simulator and the environment (Table 3). Additionally a detailed description of the simulator, the simulation event, the interaction with the simulator and the feedback provided to the participant are required. It is important that the sources of data are clearly described including when/if there are multiple sources that will be used (eg. video, checklist, simulator software).

Table 3 in the main manuscript provides a detailed description of all the key elements to report for simulation-based research. These elements can be applied to randomized trials or observational studies, and should be reported for both the intervention and control groups for randomized trials. Depending on various factors, including the type of simulation study (ie. simulation as an educational intervention vs. simulation as investigative methodology), objectives, type of simulator, and study environment (amongst others), there may be some elements that are not applicable. For example, a study assessing the value of a simulation-based intervention that does not include a feedback or debriefing component would not require reporting for the feedback and/or debriefing element and sub-elements. Similarly, a study assessing the impact of a new defibrillator on team performance during simulated cardiac arrest (ie. simulation as investigative methodology) would not require a description of instructional design elements. However, this study would require a very detailed description of participant orientation, the simulation environment, and simulator type. We encourage readers to review the entire list of key elements when writing research protocols, and referring back to the list when preparing the manuscripts for submission. A detailed description of these elements for simulation studies will allow readers to better comprehend how the various different aspects of simulation were used and controlled to address the research question.

Item: Methods/Outcomes
CONSORT: 6a: Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed; 6b: Any changes to trial outcomes after the trial commenced, with reasons.
STROBE: N/A

Extension: In describing the details of methods of assessment, include (when applicable) the setting, instrument, simulator type, timing in relation to the intervention, along with any methods used to enhance the quality of measurements. Provide evidence to support the validity and reliability of assessment tools in this context (if available).

Example
CONSORT:
"The primary outcome of the study was the time for intubation, defined as the time from the insertion of the laryngoscope blade between the teeth to the first manual ventilation of the manikin's lungs. The secondary outcome was the success of the intubation attempt (ie, tracheal or oesophageal placement of the tube), which was recorded when the success of the ventilation attempt could be seen by the manikin's ventilation indicators. Failed intubation was defined as either esophageal intubation or exceeding a time limit of 60 seconds. After each attempt, participants were asked to rate the glottic view they had during the attempt using a Cormack-Lehane grade. All processes were recorded by camera, and each time variable was precisely identified by reviewing the video recorded. We also measured dental compression, which was assessed using a visual scale grading; the pressure was applied on the upper teeth (n, none; mild, 1; moderate, 2; and severe, 3). To access a subjective opinion about the difficulty of each intubation method,

participants were asked to rate it on a visual analog scale (VAS) with a score from 1 (extremely easy) to 5 (extremely difficult). For comparisons of VAS, a Cormack-Lehane grade 1-way analysis of variance, with post hoc (Scheffé) test, was used."[11]

Explanation

It is important to clearly define all primary and secondary endpoints of study in adequate detail to allow other researchers to utilize the same outcome and measurement strategy. In simulation-based research, attention should be paid to describing the outcomes, and methods of measurement that are unique to simulation. In the examples above, there is a clear description of how the outcome was defined in the context of the manikin being used. The manikin make and model were both enumerated and there was a clear description of the setting, including the lighting in the space (which may impact outcome assessment-particularly since video was used in this case). Further description of the video, including standardization of camera angles, and how many views were obtained, is also desirable. Specifically noting whether the environment being used is a clinical space (in situ) or a simulation center can also help with generalizability of results. It is important to use validated assessment tools and the above article uses the published Cormack-Lehane grading system. Researchers should consider providing additional evidence to support the validity and reliability of an assessment tool in the context of their study. As simulation research is relatively new, many of the assessment tools or scales used in simulation research are borrowed or adapted from other disciplines and may not perform the same when implemented in a simulation context. Therefore, it is vital to describe both the existing evidence for validity and reliability and the context in which the assessment tool has previously been evaluated. Authors should also describe whether assessments are being performed live (i.e. assessment while the simulation is occurring) or after the fact (i.e. through video review), whether a live assessor is in the same room as the simulation or concealed from the trainee (i.e. behind one-way glass or via video link) and whether the assessment tool(s) have previously been used in this fashion. Cook et al described a modern validity framework that provides a consistent approach to validity and reliability[12]. The component sources of evidence include content, response process, internal structure, relations to other variables and consequences. Ideally research reports would provide evidence for all of these elements of the validity framework, however that may not always be available. In the above examples the authors provide evidence for some of these elements, however they do not report on the internal structure (reliability and factor structure). Reporting these data would require an additional research team member to score the videos, however it would increase the generalizability and strengthen their findings. Lastly, the authenticity of the simulation itself should be reported if the outcome is being measured in the simulated environment. We encourage the use of the key elements described in Table 3 as a guide to describing the context of the outcomes assessment, with specific focus on the simulator type, the simulation environment and the simulation event/scenario.

Item: Data sources / measurement

CONSORT: N/A

STROBE: For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.

Extension: In describing the details of methods of assessment, include (when applicable) the setting, instrument, simulator type, timing in relation to the intervention, along with any methods used to enhance the quality of measurements. Provide evidence to support the validity and reliability of assessment tools in this context (if available).

Examples

STROBE:

"The expert facilitators were … three pediatric critical care attending physicians, one neonatology fellow, and one PICU respiratory therapist educator. The experts rated the team performance immediately after each training session on site, while RAs rated approximately 20% of the performance via video review. Once inter-rater reliability statistics were established, actual PICU intubations were observed by one of the two trained raters (RAs) with JIT-PAPPS version 3. The training room was configured to be identical to the patient rooms. Just-in-Time training consisted of two parts: 1) hands-on training for the incoming, on-call resident with bag and mask ventilation skills and orotracheal intubation skills by using a pediatric (5-yr-old) manikin (MegaCode Kid, Laerdal, Wappingers Falls, NY), and 2) brief, standardized scenario-

based team training with a high-fidelity human infant (6 – 8 months) simulation manikin (SimBaby, Laerdal, Stavanger, Norway) in respiratory failure, which requires bag and mask ventilation and orotracheal intubation, followed by a scripted short debrief using JIT-PAPPS as a checklist."[13]

Explanation
Many simulation studies require the use of assessment tools in order to measure performance or behavior of participants. The example above enumerates steps taken by the team to develop and assess such a tool for their study, underscoring the importance of a detailed description of any assessment tool used for measuring outcomes in SBR. Measurement errors can significantly bias research results, and as such, reporting the validity and reliability of any assessment tools used for the purposes of research will help to provide clarity and context for the reader. If video recording was used for purpose of data analysis, it should be clearly noted when the analysis occurred in proximity to the intervention, whether pausing/rewinding of the video was allowed and full descriptions of the angles and views recorded. For example, when describing the manikin, or virtual programs that were used, make, models and version histories of software should be reported. As noted above, the authenticity of the simulation itself is a complex variable to be considered when outcomes are being measured in the simulated environment. Researchers should use Table 3 as a guide to describe the context of the assessment, with specific focus on the simulator type, the simulation environment and the simulation event/scenario. Describing the context of assessment will help readers to interpret results and researchers to reproduce the environment in which the study was conducted.

Items: Blinding/masking
CONSORT: 11a: If done, who was blinded after assignments to interventions (for example, participants, care providers, those assessing outcomes) and how; 11b: If relevant, description of the similarity of interventions
STROBE: N/A

Extension: Describe strategies to decrease risk of bias, when blinding is not possible.

Examples:
CONSORT:
"Study participants were randomly assigned to receive one of three on-line training sets. Concealed treatment allocation was provided by a randomization program developed using LabVIEW software. From each member institution, candidates were block randomized to the three training sets, thereby limiting any bias attributable to differences between institutions. Once randomized, participants received their respective set of 50 cases in a random order. This was immediately followed by the standardized post-test of 20 radiographs that was common to all participants. Participants were unaware of which training set they completed."[14]

Explanation
Blinding is a technique to diminish the influence of bias due to the study members' prior beliefs about the study results and is perhaps most widely known for its use in "triple-blind" placebo-controlled pharmaceutical studies where neither the study participant, nor the investigators who might determine outcome measures, nor the data analysts know which study group has received which treatment until all analyses have been completed. In the radiology simulation example described above, concealment of treatment allocation kept the participants from knowing the study group to which they had been assigned. Outcomes assessment is recognized as a significant source of bias in clinical research, and blinding the assessors can help to mitigate this risk, as can using metrics recorded automatically by the simulator. Finally, statistical analyses were performed without the analyst's knowing which group was which – this recognizes that the myriad decisions made by the analyst, including which observations to drop and which statistical tests to use, are often subjective and potentially open to bias.

Simulation-based research can leverage a variety of techniques for blinding. The name of the hospital and/or staff members can be removed from the area during a simulation in a center. These can include the use of video review techniques that blur out faces and/or remove audio, having team members wear protective equipment so their faces cannot be identified or having an individual from another center review

the videos. All of these techniques could help to diminish the influence of bias in some way, however they have their limitations. Additional techniques could utilize the simulator to capture data directly, thus removing the potential for human bias.

Item: Statistical Methods
CONSORT: Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses
STROBE: (a) Describe all statistical methods, including those used to control for confounding; (b) Describe any methods used to examine subgroups and interactions; (c) Explain how missing data were addressed; (d) Cohort study: If applicable, explain how loss to follow-up was addressed. Case–control study: If applicable, explain how matching of cases and controls was addressed. Cross-sectional study: If applicable, describe analytical methods taking account of sampling strategy; (e) Describe any sensitivity analyses.

Extension: Clearly indicate the unit of analysis (e.g. individual, team, system) and identify repeated measures on subjects, and describe how these issues were addressed.

Example
CONSORT:
"Three outcome measures were utilized: a MCQ test to assess the medical knowledge of individual participants, the Clinical Performance Tool (CPT) to assess the clinical management of the team, and the Behavioral Assessment Tool (BAT) to assess the team leader's behavioral performance. Existing evidence suggest that measures of knowledge, clinical performance and behavioral performance may be related to changes in patient care and/or outcomes **…** All data analysis was performed using statistical software JMP 7.0.1, with significance designated as a p-value < 0.05. Pearson's chi-square was used to assess if demographics were evenly distributed across study arms. Post vs. pre-simulation comparisons (PPC) scores for MCQ, BAT and CPT were calculated in percentages. As each score represents an individual or team compared to itself, this is a form of repeated measures analysis, with the advantage of limiting inter-subject variability. Shapiro-Wilk tests were used to test for normality. MCQ data were normally distributed, so means with 95% confidence intervals were reported and one-way ANOVA tested for differences between the 4 study arms. Two-sample independent t-tests (performed on individual post-pre scores) queried for differences between scripted vs. non-scripted MCQ-PPC and HiR vs. LoR MCQ-PPC. As BAT and CPT data were not normally distributed, medians with IQRs were reported, and Kruskal–Wallis one-way analysis of variance test was used. Mann-Whitney tests queried for differences between scripted vs. non-scripted BAT and CPT-PPC scores and HiR vs. LoR BAT and CPT-PPC.**"**[9]

STROBE:
"Data collection included the number of mock codes conducted, reported learning outcomes from the mock code events, residents' self-perception ratings, and real CPA survival rates for pediatric patients. We asked residents to "Please list what you learned during today's mock code" to assess their perceived learning outcomes and used a 6-point rating scale (1 = very poor; 6 = outstanding) to capture residents' self-perceptions of their ability to lead an actual code. The University of Michigan Hospitals and Health Centers Office of Clinical Affairs provided hospital records for pediatric resuscitation survival rates for 48 months after its start (2005–2008) ... Quantitative and qualitative methods were used to analyze the resulting data. We calculated effects sizes (Cohen's *d*), and used SPSS 16.0 to calculate descriptive and inferential statistics with statistical significance set at $p < 0.05$ and SPSS Text Analysis for Surveys 3.0 for qualitative analyses of reported learning outcomes (SPSS Inc., Chicago, IL) …We calculated the mean and SD values for residents' perceptions of their ability to lead an actual code."[15]

Explanation
In a simulation-based study a variety of levels of data can be used for analysis. It is important for investigators to describe their primary outcome and the level of analysis that they are exploring. Simulation-based studies will involve examining impact at the level of the provider (eg. through a test or performance checklist, teams of providers (eg. through a clinical scoring tool) or at the level of the system (eg. frequency of adverse errors). Simulation-based researchers can also analyze outcomes across a variety of settings. In a series of studies investigating central line placement, Barsuk et al have examined

performance at the level of the provider in the simulation laboratory (T1)[16], at the level of the provider in the clinical environment (T2)[16], and at the level of the patient by reporting central line infection rates[17]. In studies of team function, it is important to clarify if the randomization and analysis are to be conducted at the level of the individual, the team or even a higher level, such as a course or an entire clinical unit or hospital and how many of each type of subject are enrolled[18].

Item: Baseline Data/Descriptive Data
CONSORT: A table showing baseline demographic and clinical characteristics of each group
STROBE: (a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders; (b) Indicate the number of participants with missing data for each variable of interest; (c) Cohort study: Summarize follow-up time—e.g., average and total amount

Extension: In describing characteristics of study participants, include their prior experience with simulation and other relevant features as related to the intervention(s).

Examples
CONSORT:

Table 1. Demographic Characteristics of Team Leaders and CPR Providers

Characteristics	Team Leader N=108	CPR Providers N=216
	N(%)	
Gender		
Male	53(49.1)	52 (24.1)
Female	55 (50.9)	164 (75.9)
Occupation (Level of Training)		
Attending physician	16 (14.8)	4 (1.9)
Resident	92 (85.2)	63 (29.2)
Medical Student	0 (0)	26 (12.0)
Nurse	0 (0)	120 (55.6)
Other (RT, paramedics)	0 (0)	3 (1.4)
Last BLS course taken		
Never	1 (0.9)	5 (2.3)
< 12 month	51 (47.2)	142 (65.7)
> 12 month	56 (51.9)	69 (31.9)
Last ACLS course taken		
Never	35 (32.4)	107 (49.5)
< 12 month	20 (18.5)	49 (22.7)
> 12 month	53 (49.1)	60 (27.8)
Last PALS course taken		
Never	2 (1.9)	68 (31.5)
< 12 month	65 (60.2)	90 (41.7)
> 12 month	41 (38.0)	58 (26.9)
Chest compression on pediatric patient within 2 years		
Never	39 (36.1)	134 (62.0)
1-5 times	51 (47.2)	69 (31.9)

6 times or more	18 (16.7)	13 (6.0)
Chest compression on manikin within 2 years		
Never	7 (6.5)	53 (24.5)
1-5 times	67 (62.0)	123 (56.9)
6 times or more	34 (31.5)	40 (18.5)
Total Participants	108 (100.0)	216 (100.0)

Adapted from Table 1 of Cheng A et al. Published with permission, Copyright Clearance Center (Elsevier)[19]

STROBE:

Table 1. Baseline characteristics

Baseline Characteristics	Pre-RCDP, *n* = 70 *n* (%)	Post-RCDP, *n* = 51 *n* (%)	*p*-Value
PGY			
1	27 (39%)	19 (37%)	0.975
2	22 (31%)	17 (33%)	
3	21 (30%)	15 (29%)	
Gender			
Males	28 (40%)	19 (37%)	0.760
Females	42 (60%)	32 (63%)	
Resuscitation training in medical school			
BLS	57 (81%)	41 (82%)	0.936
ACLS	33 (47%)	14 (28%)	0.034
PALS	3 (4%)	2 (4%)	0.938
Code team training	23 (33%)	12 (24%)	0.293
Simulator exposure	17 (24%)	13 (26%)	0.831
Resuscitation using simulator	8 (11%)	10 (50%)	<0.001
Resuscitation training in residency			
Mock code this year	46 (66%)	37 (79%)	0.129
Mock code during residency	65 (93%)	46 (92%)	0.860
Simulator training during test year	37 (53%)	22 (45%)	0.393
Resuscitation training ever			
BLS	64 (91%)	44 (94%)	0.663
PALS	40 (57%)	46 (98%)	<0.001
Resuscitation experience			
Attended real code in medical school	53 (76%)	37 (73%)	0.694
# codes in medical school[a]	2 (1–3)	1 (0–2)	0.2767
Attended real code in residency	67 (96%)	45 (88%)	0.122
# codes in residency[a]	4 (3–8)	5 (2–7)	0.4208
Codes as Team Leader	0 (0–0)	0 (0–0)	0.6125
Ever defibrillated a patient	11 (16%)	6 (12%)	0.595
Ever defibrillated a patient or manikin	46 (66%)	41 (80%)	0.076

Adapted from Table 1 of Hunt EA et al. Published with permission, Copyright Clearance Center (Elsevier)[20]

"Baseline characteristics are stratified by cohort and reported in Table 1. In the table resuscitation training was similar between the two cohorts except fewer of the post-intervention cohort had training in advanced cardiopulmonary life support and more had exposure to simulators during medical school. As expected, more post-intervention than pre-intervention residents had taken PALS since one of the recommendations from the initial study was to move PALS into the internship year."[20]

Explanation
Simulation has been adopted as a training modality across health care institutions in various forms. Reporting experience with simulation in general, simulation related to focus of the study, and timing and frequency of simulation experience will provide insight into the degree of exposure to simulation for the participants enrolled in the study. Participants who are highly versed with, and familiar with the nuances of performing in the simulated environment (in general, or specific to the focus of the study) may have an added advantage, thus potentially introducing a threat to the internal validity of the study. Reporting the previous simulation experience of participants provides clarity for readers, reviews and editors who are assessing the strengths and weaknesses of the study, and trying to determine if the assertions made by the author regarding the effectiveness of interventions are truly valid and/or generalizable.

Item: Outcomes and Estimation/Main Results
CONSORT: 17a: For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval); 17b: For binary outcomes, presentation of both absolute and relative effect sizes is recommended
STROBE: (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included; (b) Report category boundaries when continuous variables were categorized; (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.

Extension: For assessments involving more than one rater, inter-rater reliability should be reported.

Examples
CONSORT:
"All pre- and posttests were graded on the checklist by a single unblinded instructor (J.H.B.) and were videotaped. If the artery was punctured, or >2 needle passes were made before cannulating the vein, the test was stopped and the remaining steps were marked wrong. A 50% random sample of pre and posttest sessions were rescored by a second rate to assess inter-rater reliability. The second rater was blind to the results of the first checklist recording and to the pre or posttest status of the examinee. Checklist score reliability was estimated by calculating inter-rater reliability, using the [kappa] coefficient adjusted according to the formula of Brennan and Prediger. Inter-rater reliability measured by the mean [kappa] coefficient was very high ($[kappa]_n = 0.91$) across the 27 IJ and SC checklist items."[21]

STROBE:
"Minimally interrupted cardiac resuscitation protocol compliance was determined by all 4 components: 200 pre-shock chest compressions, 200 post-shock chest compressions, delayed intubation attempt for 3 cycles of 200 compressions and rhythm analysis, and patients having received intravenous epinephrine in the first or second cycle of chest compressions."[22]

Table 2: Final Logistic Regression Models for Survival and Neurologic Outcomes (all rhythms and witnessed/shockable)

Survival to Hospital Discharge and Associated OR (95% CI)				
Characteristic	**No./Total**	**Absolute Difference**	**Crude OR (95% CI)**	**Adjusted OR (95%)***
Study Period Pre Post	 20/231 (8.7) 35/252	 5.2 (-0.4 to 10.8)	 1 (Reference) 1.73 (0.93 to	 1 2.72 (1.15 to

	(13.9)		3.21)	6.41)
Witness Arrest No Yes	 13/291 (4.5) 42/192 (21.9)	 17.4 (11.1 to 23.7)	 1 4.19 (2.12 to 8.3)	 1 4.00 1.72 to 9.28)
Initial Rhythm Nonshockable VF/VT	 12/334 (3.6) 43/149 (28.9)	 25.3 (17.7 to 32.8)	 1 7.33 (3.59 to 14.99)	 1 5.88 (2.59 to 13.36)
Use of TH No Yes Age, per year	 25/431 (5.8) 30/52 (57.7)	 51.9 (38.3 to 65.5)	 1 14.60 (5.16 to 27.33) 0.97 (0.95 to 0.99)	 1 11.87 (5.16 to 27.33) 0.98 (0.95 to 1.00)
MICR protocol compliance Partial Complete	 10/108 (9.3) 45/375 (12.0)	 2.7 (-3.6 to 9.1)	 1 1.16 (0.46 to 2.93)	 1 1.16 (0.46 to 2.93)
VF: ventricular fibrillation; VT: ventricular tachycardia; MICR: minimally interrupted cardiac resuscitation * Adjusted for all variables listed in this table in final model (likelihood ratio P value for all variables included in final model ≤ .05 or judged a significant confounder).				

Adapted from Table 2 of Bubrow BJ et al. Published with permission, Copyright Clearance Center (Elsevier)

Explanation

Using raters for performance assessment is fairly common in quantitative simulation-based research. When more than one rater is used, the inter-rater reliability (IRR) needs to be measured and reported. IRR provides a clear estimation of degree of consistency or reliability among the raters, and also allows researchers and readers to gauge the strength of evidence collected through the assessment of multiple raters. Researchers who conduct pilot testing of their assessment tools may encounter a suboptimal IRR during the early phase of simulation-based study. This represents a potential issue that may exist within the study methodology, particularly in the data collection processes such as the lack of reliability in scoring criterions, inconsistency of camera recording angles, or lack of standardization of the rater training. Therefore, measuring IRR provides advantages at different stages of the study. Authors should be sure to report details about how the calculation was performed, and also the number of raters reflected in the reported reliability estimate (e.g., in a study with 3 raters, the IRR would usually be reported as if only 1 rater had provided ratings [single-rater reliability], but occasionally authors report the reliability for all 3 raters which will inevitably be higher).

Item: Discussion/Limitations

CONSORT: Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses

STROBE: Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.

Extension: Specifically discuss the limitations of simulation-based research.

Examples

CONSORT:

"Our study did not demonstrate the same benefits of a high realism (e.g. turned "on") simulator. The effect may have been diminished because other aspects of realism in our study were high. All participants in the

study were exposed to a simulated environment with very high physical realism. The pre-programming of scenarios helped to ensure a high degree of semantical realism. In the low realism groups, phenomenal realism was optimized by using facilitator guided verbal cues at pre-defined times. Additionally, the scenario we selected for the study involved a pulseless patient, thus not demanding much physical feedback from the mannequin to simulate reality. Thus, our findings related to the lack of simulator physical realism are likely tempered by relatively high degrees of physical, semantical and phenomenal realism, and possibly by selection of a scenario with limited physical findings."[9]

STROBE:
"Our study had several key limitations… Another limitation of our study was the use of a low- fidelity mannequin. The mannequin could not show signs of clinical improvement or deterioration and required our study personnel to provide verbal prompts of clinical status changes. It is possible that the ED teams would have performed more effectively on a real child or on a higher fidelity mannequin. However, young children and critically ill patients are often unable to cooperate and answer questions appropriately at the time of stabilization or resuscitation event, making it all the more important for providers to complete every component of the primary and secondary surveys suggested by ATLS guidelines."[23]

Explanation
Simulation-based research has several inherent limitations that should be reported and described in the context of the study objective(s). The most obvious limitation relates to the issue of realism or authenticity. When the simulated patient and/or simulated clinical environment is not authentic enough to replicate the similar experience in real life, subjects may alter their behavior or their performance may not be reflective of that in the real clinical environment[8]. For example, in a study assessing the quality of chest compressions, an adult manikin that does not allow compressions to beyond the recommended depth (ie. 5 cm) would introduce a significant confounding variable to the study. Alternatively, the results of a study assessing team performance during an obstetrical emergency conducted in the simulation lab may differ substantially from the same study conducted in an in situ environment. Researchers should be transparent in reporting how the limitations of the simulated patient and/or the simulated environment affect the outcomes of the study.

Item: Discussion/Generalizability
CONSORT: Generalizability (external validity) of the trial findings
STROBE: Discuss the generalizability (external validity) of the study results.

Extension: Describe generalizability of simulation-based outcomes to patient-based outcomes (if applicable).

Examples
CONSORT:
"Despite the attention to multi-center standardization, our study had several limitations. We had one simulated CPA scenario with 2 team members providing CPR. As such, the results of our study reflect this specific simulated context, making it difficult to predict if our results are directly generalizable to real patient care. We had 45 minutes between JIT CPR training and the CPA scenario, while in a real clinical environment the delay between JIT CPR training and provision of CPR is unpredictable. Furthermore, participants in the JIT CPR training groups only received one JIT session. The eventual quality of CPR provided was likely influenced by these factors (ie. time delay between JIT training and CPR event; frequency of JIT training). We excluded participants who had previous experience with CPR feedback devices from our study to remove this as a confounding variable. As such, the effect of the identical interventions on providers with previous experience using CPR feedback devices was not assessed. We believe the results of our study are generalizable to similar healthcare providers naïve to CPR feedback devices across a broad spectrum of geographic locations."[24]

STROBE:
"A final limitation of our study is that we have not determined if an improvement in simulation performance by ED teams will translate into improved pediatric trauma patient outcomes. Participation in simulated exercises may help clinicians improve their management of low-occurrence, high-risk emergency

medical situations such as cardiac arrests and trauma stabilization … we should take advantage of opportunities to practice on plastic first, in an effort to increase our clinical competency and to enhance patient safety. While a growing body of simulation research supports this hypothesis, future studies are needed to define its effect."[23]

Explanation
The applicability and transferability of findings from the simulated environment to outcomes in the real world setting (generalizability) is a vital consideration[25,26]. Whilst much attention in this document has been appropriately focused on the importance of controlling for confounding variables that threaten internal validity of simulation based studies[8], it is equally important to recognize that tightly controlling and standardizing the simulated environment may limit the generalizability of findings to the real clinical environment.

By tightly controlling the environment for simulation research, the researcher may, at the same time, be reducing the generalizability of their findings. For example, in a study assessing the value of simulation-based team training for emergency room providers, the researcher decides to use only one cardiac arrest scenario during the study as the context for training. The scenario is tightly scripted to ensure that it is conducted in a uniform manner from one group to the next. While this strategy helps to reduce simulation-specific threats to internal validity, the simple fact that the training was done in a limited context reduces the generalizability of the study. Studies utilizing low or high fidelity manikins may also suffer from the same issues. While use of one single manikin for training and/or assessment in the context of a research study is desirable from a standardization point of view, the results of the study may be difficult to generalize to real patients whose physical characteristics are variable from one patient to the next. In some select instances, previous studies have demonstrated an association between improved performance in the simulated environment with improved clinical outcomes[26]. If this evidence exists for a specific outcome of interest, this should be described and/or cited in the discussion section of the manuscript.

One strategy to improve generalizability is to take a random sampling of a study sample related to the general population of interest. Simulation-based research typically involves health care providers serving as the subjects. Frequently the subjects are collected non-randomly due to different pre-existing constraints (e.g. a group of nursing students from a particular class or a group of physician trainees from a particular training program). Limited sample representativeness constraints the generalizability of findings to and across the general population of interest. For example, the efficacy of a simulation-based training program in a nursing population from a particular university may be reasonably different when applied to a different population of physicians in training, even if the simulation-based training program is to be replicated exactly. It is important to be aware of this potential issue for both the authors and readers in discussing the generalizability of study findings.

Item: Discussion/Funding
CONSORT: Sources of funding and other support (such as supply of drugs), role of funders
STROBE: Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.

Extension: List simulator brand and if conflict of interest for intellectual property exists.

Example
CONSORT / STROBE:
"Conflict of Interest Disclosures: … reports receiving a research grant from the American Heart Association (AHA) for design and conduct of the study and collection and analysis of data, a grant from the Laerdal Foundation for Acute Medicine, and an infrastructure grant for the EXPRESS collaborative to support administrative and technical positions (funds from the Laerdal Foundation for Acute Medicine grant were not used for conducting this study) … reports receiving a research grant from the Laerdal Foundation for Acute Medicine, The Hartwell Foundation, and money paid for expert testimony by DeBlasio&Donnell LLC…
Funding/Support: This study was funded by an educational research grant from the AHA.
Role of the Sponsors: Funds from this grant were used for the design and conduct of the study, as well as

collection, management, analysis, and interpretation of data."[9]

Explanation

Funding and conflicts of interest are explicitly detailed to allow the reader to critically appraise the potential threats to the internal validity of the study. This is to include the simulator company and/or manufacturer providing support (if applicable), and if any of the investigators have a potential conflicts of interest related to the simulator used and/or the simulation company. Inclusion of all funding streams for the study allows the reader to make an assessment of potential bias. An explicit statement highlighting the role of the sponsors for the study, and providing detail to the reader of how funding was utilized at each stage of the study should be included.

References

1. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; 340:c869.
2. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al., for the STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007; 4:e297
3. Moak JH, Larese SR, Riordan JP, Sudhir A, Yan G. Training in transvaginal sonography using pelvic ultrasound simulators versus live models: a randomized controlled trial. *Acad Med*. 2014; 89:1063-1068.
4. Wayne DB, Cohen ER, Singer BD et al. Progress toward improving medical school graduates' skills via a "boot camp" curriculum. *Simul Healthc.* 2014; 9:33-39.
5. National Center for Biotechnology Information. PubMed. Patient Simulation. http://www.ncbi.nlm.nih.gov/mesh/?term=%22patient+simulation%22. Accessed on May 28, 2015.
6. Siegel NA, Kobayashi L, Dunbar-Viveiros JA et al. In situ medical simulation investigation of emergency department procedural sedation with a randomized trial of experimental bedside clinical process guidance intervention [published online April 3, 2015]. *Simul Healthc*. doi: 10.1097/SIH.0000000000000083
7. Kessler D, Pusic M, Chang TP et al. Impact of Just-in-Time and Just-in-Place simulation on intern success with infant lumbar puncture. *Pediatrics*. 2015; 135:e1237-1246.
8. Cheng A, Auerbach M, Chang T, Hunt EA, Pusic M, Nadkarni V, Kessler D. Designing and Conducting Simulation-based Research. *Pediatrics.* 2014; 133:1091-1101.
9. Cheng A, Hunt EA, Donoghue A, et al. Examining Pediatric Resuscitation Education Using Simulation and Scripting (EXPRESS): A Multicenter, Randomized-Controlled Trial. *JAMA Pediatrics*. 2013. 167:528-36.
10. Dissertations & Theses – Gradworks. An evaluation of the quality of resuscitation delivered during simulated pediatric in-hospital cardiopulmonary arrests, by Hunt, Elizabeth A, PhD. http://gradworks.umi.com/33/09/3309820.html. Accessed May 28, 2015.
11. Szarpak L, Karczewska K, Czyzewski L, Kurowski A. A randomized comparison of the Laryngoscope with Fiber Optic Reusable Flexible Tip English Macintosh blade to the conventional Macintosh laryngoscope for intubation in simulated easy and difficult child airway with chest compression scenarios [published online April 15, 2015]. *Am J Emerg Med*. doi: 10.1016/j.ajem.2015.04.015.
12. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006; 119:116.e7-16.
13. Nishisaki A, Nguyen J, Colborn S, et al. Evaluation of multidisciplinary simulation training on clinical performance and team behavior during tracheal intubation procedures in a pediatric intensive care unit. *Pediatric Critical Care Medicine*. 2011; 12: 406-414.
14. Pusic M, Andrews J, Kessler D, et al. Prevalence of Abnormals In An Image Bank Affects the Learning of Radiograph Interpretation. *Medical Education*. 2012; 46: 289-98.
15. Andreatta P, Ernest S, Thompson M, Annich G. Simulation-based mock codes significantly correlate with improved pediatric patient cardiopulmonary arrest survival rates. *Pediatr Crit Care Med*. 2011; 12: 33-38.
16. Barsuk JH, McGaghie WC, Cohen ER, Balachandran JS, Wayne DB. Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit. *J Hosp Med*. 2009; 4:397-403.
17. Barsuk JH, Cohen ER, Potts S et al. Dissemination of a simulation-based mastery learning intervention reduces central line associated bloodstream infections. *BMJ Qual Saf.* 2014; 23:749-756.
18. Hunt EA, Cruz-Eng H, Bradshaw JH, et al. A novel approach to life support training using "action-linked phrases". *Resuscitation*. 2014;86:1-5.
19. Cheng A, Overly F, Kessler D, et al. Perception of CPR Quality: Influence of CPR Feedback, Just-in-Time Training and Provider Role. *Resuscitation*. 2015; 87: 44-50

20. Hunt EA, Duval-Arnould JM, Nelson-McMillan KL, Bradshaw JH, Diener-West M, Perretta JS, Shilkofski NA. Pediatric resident resuscitation skills improve after "rapid cycle deliberate practice" training. *Resuscitation*. 2014; 85:945-51.
21. Barsuk JH, McGaghie WC, Cohen ER, O'Leary KJ, Wayne DB. Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med.* 2009; 37:2697-2701.
22. Bobrow BJ, Vadeboncoeur TF, Stolz U, et al. The influence of scenario-based training and real-time audiovisual feedback on out-of-hospital cardiopulmonary resuscitation quality and survival from out-of-hospital cardiac arrest. *Ann Emerg Med*. 2013; 62:47-56.e1
23. Hunt EA, Heine M, Hohenhaus SM, Luo X, Frush KS. Simulated pediatric trauma team management: assessment of an educational intervention. *Ped Emerg Care*. 2007; 23:796-804.
24. Cheng A, Brown L, Duff J, et al. Improving CardioPulmonary Resuscitation with a CPR Feedback Device and Refresher Simulations (CPR CARES Study): A Multicenter, Randomized Trial. *JAMA Pediatrics*. 2015; 169(2):137-144.
25. McGaghie WC, Draycott TJ, Dunn WF, Lopez CM, Stefanidis, D. Evaluating the Impact of Simulation on Translational Patient Outcomes *Simul Healthc*. 2011; 6: S42–S47.
26. Brydges R, Hatala R, Zendejas B, Erwin PJ, Cook DA. Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Acad Med.* 2015; 90:246-256.