UCLA
UCLA Previously Published Works

Title
Critical appraisal of emergency medicine education research: The best publications of 2012

Permalink
https://escholarship.org/uc/item/2nz1586q

Journal
Academic Emergency Medicine, 21(3)

ISSN
1069-6563

Authors
Lin, M
Fisher, J
Coates, WC
et al.

Publication Date
2014

DOI
10.1111/acem.12336

Peer reviewed

EDUCATIONAL ADVANCE

Critical Appraisal of Emergency Medicine Education Research: The Best Publications of 2012

Michelle Lin, MD, Jonathan Fisher, MD, MPH, Wendy C. Coates, MD, Susan E. Farrell, MD, EdM, Philip Shayne, MD, Lauren Maggio, MS, MA, and Gloria Kuhn, DO, PhD

Abstract

Objectives: The objective was to critically appraise and highlight medical education research published in 2012 that was methodologically superior and whose outcomes were pertinent to teaching and education in emergency medicine (EM).

Methods: A search of the English language literature in 2012 querying Education Resources Information Center (ERIC), PsychInfo, PubMed, and Scopus identified EM studies using hypothesis-testing or observational investigations of educational interventions. Two reviewers independently screened all of the publications and removed articles using established exclusion criteria. This year, publications limited to a single-site survey design that measured satisfaction or self-assessment on unvalidated instruments were not formally reviewed. Six reviewers then independently ranked all remaining publications using one of two scoring systems depending on whether the study methodology was primarily qualitative or quantitative. Each scoring system had nine criteria, including four related to methodology, that were chosen a priori, to standardize evaluation by reviewers. The quantitative study scoring system was used previously to appraise medical education published annually in 2008 through 2011, while a separate, new qualitative study scoring system was derived and implemented consisting of parallel metrics.

Results: Forty-eight medical education research papers met the a priori criteria for inclusion, and 33 (30 quantitative and three qualitative studies) were reviewed. Seven quantitative and two qualitative studies met the criteria for inclusion as exemplary and are summarized in this article.

Conclusions: This critical appraisal series aims to promote superior education research by reviewing and highlighting nine of the 48 major education research studies with relevance to EM published in 2012. Current trends and common methodologic pitfalls in the 2012 papers are noted.

ACADEMIC EMERGENCY MEDICINE 2014; 21:322–333 © 2014 by the Society for Academic Emergency Medicine

Quality, hypothesis-driven education research is necessary to promote evidence-based decisions about effective ways to teach the physicians of tomorrow. Education research has gained increasing support and prominence in emergency medicine (EM) academia with available grant opportunities from the Society for Academic Emergency Medicine (SAEM) and the Council of Emergency Medicine Residency Directors (CORD). Furthermore, the 2012 *Academic Emergency Medicine* consensus conference focused on the theme

"Education Research in Emergency Medicine: Opportunities, Challenges, and Strategies for Success" to promote a national initiative to advance the field of education research.[1]

In this fifth installment of the annual critical appraisal series, the same six reviewers used previously published criteria to critically analyze and rank the EM education research from 2012. The focus of this article is to review and highlight the methodologically superior studies that are pertinent to teaching and education in EM. Trends in EM education research over the past 5 years are summarized. We hope that this paper will serve as a valuable resource for EM educators and researchers invested in the scholarship of teaching.[2]

METHODS

Article Identification
A medical librarian (LM) performed the literature search in the medical and social sciences literature domains and supplied medical subject heading (MeSH) and keyword terms. MEDLINE was searched through PubMed using a Boolean search strategy that incorporated the following MeSH terms: emergency medicine and medical education, medical student, internship and residency, teaching rounds, undergraduate medical education, graduate medical education, and continuing medical education. Keyword variants for the MeSH terms were included in the search for comprehensiveness. Boolean searches of other databases, including Scopus ("medical education" and emergency), Education Resources Information Center (ERIC; emergency medicine), and PsychInfo (emergency medicine and education) were performed using keyword searching and where possible using the databases' controlled vocabularies. Publications were limited to English language papers published in 2012. Searches were performed in December 2012, January 2013, and February 2013.

Inclusion and Exclusion Criteria
We only included medical education studies that enrolled EM learners (students, trainees, or attending physicians) or EM educators. Medical education studies were defined as hypothesis-testing investigations and measurements of educational interventions using either quantitative or qualitative methods. Publications were excluded if they were opinion, comments, literature reviews, descriptive, or reports on education of prehospital personnel or if they could not be generalized to EM training outside of the countries in which they were performed.

Data Collection
Two authors independently screened abstracts of all retrieved publications and applied the exclusion criteria. All differences in opinion were resolved by discussion. Retrieved publications were maintained in an EndNote X5 (Thomson Reuters, New York City, NY) database. Studies that were based on single-site surveys measuring primarily learner satisfaction or self-assessment scores using unvalidated instruments were removed from the final list of publications to be scored by all six reviewers, because these are generally regarded as the lowest level studies as outlined by the four-tiered Kirkpatrick model and usually include very small sample sizes. Higher-quality studies, according to the hierarchical model, involve the assessment of learning using such objective measures as written tests (second tier), the assessment of learner behavior using observer checklists and performance indicators (third tier), and the assessment of benefits at the organizational and patient level (fourth and highest-quality tier).[3,4] Publications from the final list were posted in a shared folder online for all six reviewers to score independently.

Scoring
The publications were first assigned to a scoring system based on whether they were primarily quantitative or qualitative studies. The quantitative studies used scoring criteria developed in 2009 based on domains from the validated Medical Education Research Study Quality Instrument (MERSQI) tool and then continually optimized and modified annually from 2010 through 2012.[5–9] This year, continued slight modifications focused on improved clarity of the subdomain descriptors. Quantitative studies were scored in nine domains for a maximum total score of 25 points. The domains included the following: Introduction (0–3 points), Measurement (0–4 points), Data Collection (0–4 points), Data Analysis (0–3 points), Discussion (0–3 points), Limitations (0–2 points), Innovation (0–2 points), Generalizability (0–2 points), and Clarity of Writing (0–2 points). Each study that conducted a power analysis was awarded an additional point in the Data Analysis domain. Each of the domains were scored based on predefined criteria to make scoring as objective as possible. The detailed scoring criteria used are seen in Table 1.

Using accepted recommendations and hierarchical formulations,[10–13] qualitative studies were assessed and scored in the same nine domains with Measurement, Data Collection, and Data Analysis criteria specific to qualitative research for a maximum total score of 25 points. Table 2 outlines the detailed, predefined scoring criteria.

Data Analysis
Reviewers were excluded from scoring publications where there was significant conflict of interest (own publication, own institution, or had a vested interest in the authors or work). Publications were listed alphabetically by first author surname, and each reviewer was assigned a different place to start on the list in an attempt to prevent bias resulting from reviewer fatigue. Each reviewer independently reviewed and rated the publications, and a total rating score was calculated for each article. All rating scores were entered into a spreadsheet using Microsoft Excel 2010 (Microsoft Inc., Redmond, WA). Using each reviewer's total rating score for each article, a rank list of quantitative studies and a rank list of qualitative studies were created for each reviewer. The rankings were then averaged among all six reviewers to prevent overvaluing any one reviewer's scoring. The a priori criteria for quantitative studies to be featured as exemplary were: 1) the average of all reviewers' rankings of an article placed the article's rank in the top 10 and 2) at least $(n - 1)$ reviewers

Table 1
EM Education Research Scoring System: Quantitative Research

Domain	Item	Item Score	Maximum Domain Score
Introduction: Give 1 point for each criterion met			3
	Appropriate description of background literature	1	
	Clearly frame the problem	1	
	Clear objective/hypothesis	1	
Measurement: Give 0 or 1 point for each criterion met			4
1. Methodology			
	Has no pretest or posttest	0	
	Has a posttest only	1	
	Has a pretest and posttest	1	
2. Groups			
	Both experimental and control group	1	
	Random assignment to groups	1	
Data Collection: Give 0 or 1 point for each criterion met			4
1. Institutions			
	Single institution	0	
	At least two institutions	1	
	More than two institutions	1	
2. Response rate			
	Response rate < 50% or not reported	0	
	Response rate ≥ 50%	1	
	Response rate ≥ 75%	1	
Data Analysis: Give 0 or 1 point for each criterion met			3
1. Appropriateness			
	Data analysis inappropriate for study design and type of data	0	
	Data analysis appropriate for study design and type of data	1	
2. Sophistication			
	Descriptive analysis only	0	
	Beyond descriptive analysis	1	
	Includes power analysis	1	
Discussion: Give 1 point for each criterion met			3
	Data support conclusion	1	
	Conclusion clearly addresses hypothesis/objective	1	
	Conclusions placed in context of literature	1	
Limitations: Assign a single best score			2
	Limitations not identified accurately	0	
	Some limitations identified	1	
	Limitations well addressed	2	
Innovation of Project: Assign a single best score			2
	Previously described methods	0	
	New use for known assessment	1	
	New assessment methodology	2	
Relevance of Project: Assign a single best score			2
	Impractical to most programs	0	
	Relevant to some	1	
	Highly generalizable	2	
Clarity of Writing: Assign a single best score			2
	Unsatisfactory	0	
	Fair	1	
	Excellent	2	
Total			25

ranked the article in their individual top 10 rankings, where *n* is the number of eligible reviewers. Because of the historical paucity of published qualitative studies, the single highest scoring qualitative study would be highlighted. Data were further analyzed using IBM SPSS 21.0 (Armonk, NY) for internal consistency and interrater reliability with Cronbach's alpha and intraclass correlation coefficients (ICCs) using absolute agreement, respectively.

RESULTS

A total of 564 papers satisfied the search criteria, and 48 papers met the inclusion criteria.[14–61] Fifteen papers were removed from the full-group review because they were single-site, satisfaction, or self-assessment surveys. A total of 33 articles (30 quantitative and three qualitative studies) were critically appraised by each of six reviewers, with a range of mean scores from 10.0 to 21.2 (maximum 25 points). The mean score for the 30 quantitative studies was 14.8 with a standard deviation (SD) of ±3.0, and the quantitative scoring tool demonstrated a Cronbach's alpha of 0.913 and an ICC of 0.613. Similar statistics were not calculated for the qualitative studies because there were only three included studies. Seven quantitative studies met a priori criteria as methodologically superior publications in education research.[33,35,36,41,43,46,55] Because two qualitative studies

Table 2
EM Education Research Scoring Sheet: Qualitative Research

Domain	Item	Item Score	Maximum Domain Score
Introduction: Give 1 point for each criterion met			3
	Appropriate description of background literature	1	
	Clearly frame the problem	1	
	Clear objective/hypothesis	1	
Measurement: Give 1 point for each criterion met			3
1. Methodology			
	Appropriate for study question	1	
2. Sampling of participants			
	Appropriate study population	1	
	Enrolled full range of cases/settings beyond convenience	1	
Data Collection: Give 0–1 point for each criterion met			3
1. Institutions			
	Single institution	0	
	At least two institutions	1	
	More than two institutions	1	
2. Sample size determination			
	Appropriate sample size determination	1	
Data Analysis: Give 1 point for each criterion met			5
	Clear, reproducible "audit trail" documenting systematic procedure for analysis	1	
	Data saturation through a systematic iterative process of analysis	1	
	Addressed contradictory responses	1	
	Incorporated validation strategies (e.g., member checking, triangulation)	1	
	Addressed reflexivity (impact of researcher's background, position, biases on study)	1	
Discussion: Give 1 point for each criterion met			3
	Data support conclusion	1	
	Conclusion clearly addresses hypothesis/objective	1	
	Conclusions placed in context of literature	1	
Limitations: Assign a score			2
	Limitations not identified accurately	0	
	Some limitations identified	1	
	Limitations well addressed	2	
Innovation of Project: Assign a score			2
	Previously described methods	0	
	New use for known assessment	1	
	New assessment methodology	2	
Relevance of Project: Assign a score			2
	Impractical to most programs	0	
	Relevant to some	1	
	Highly generalizable	2	
Clarity of Writing: Assign a score			2
	Unsatisfactory	0	
	Fair	1	
	Excellent	2	
Total			25

scored similarly with the highest scores of 16.8 and 17.5 (out of a maximum 25 points), both were highlighted.[19,45] The seven best quantitative studies are presented in alphabetical order by the surname of the first author, followed by the two best qualitative studies.

Harvey A, Bandiera G, Nathens AB, LeBlanc VR. Impact of stress on resident performance in simulated trauma scenarios. J Trauma Acute Care Surg 2012; 72: 497–503.[33]

Background. The effect of stress on clinical performance and training is poorly understood. This study examined whether a difference in stress response and in performance could be demonstrated between a low- and a high-stress trauma resuscitation simulation.

Methods. This was a prospective, single-blinded, case-control, crossover trial where EM and general surgery residents were evaluated during two trauma simulations,

one being the low-stress simulation in which a seriously injured patient was generally stable, and a second high-stress encounter where the patient was a persistently hypotensive, pregnant paramedic with a distraught partner, loud monitors, and discord among trauma team members. The residents were randomized to the order of the two simulations. The residents' stress response was measured by heart rate, saliva cortisol levels, a validated stress index score, and a self-assessment. Performance on the simulation was judged by two expert raters who independently viewed and scored the resuscitations using a checklist, a global rating form, time measures, and an Anesthesia Non-Technical Skills (ANTS) scale. Participants completed post-simulation recalls. An analysis suggested a minimum of 12 participants were required to attain sufficient power for significance.

Results. Thirteen residents participated. While no difference was found in heart rate, the high-stress

simulation showed a significant increase in cortisol response and the two stress indicators. Performance scores were mixed with two assessments showing worse performance for the high-stress simulation compared to the low-stress simulation and the other two assessments showing no change. Global evaluation scores were not significantly different between high-stress and low-stress simulations.

Strengths of the Study. This was a prospective, randomized, case–control study that used multiple objective measurements of stress and performance. The methodology was rigorous and the independent raters demonstrated good interrater reliability.

Relevance for Future Educational Advances. While the results were mixed, the methodology is superior and demonstrates a good model for further studies incorporating measurements of stressors on clinical performance.

Keijzers G, Sithirasenan V. The effect of a chest imaging lecture on emergency department doctors' ability to interpret chest CT images: a randomized study. Eur J Emerg Med 2012; 19:40–5.[35]
Background. The use of chest computed tomographies (CTs) in the emergency department (ED) is increasing. This study assessed emergency physician knowledge and skills of interpreting chest CT imaging before and after a 1-hour lecture.

Methods. This prospective, two-center, unblinded, randomized controlled study assessed the written test results of emergency physicians on chest CT anatomy and image interpretation. The results of the control group, who did not attend a 1-hour lecture, were compared to those of the intervention group who did attend the lecture. The outcome measures included anatomy knowledge scores, diagnosis scores, and overall scores. A power calculation was performed based on a similar study on brain CT interpretation, which determined that a sample size of 17 participants was needed for each study arm.

Results. Sixty physicians were randomized, although two physicians did not complete the study. The intervention group ($n = 27$) did not have improved knowledge (72.9% vs. 70.2%), diagnosis (71.2% vs. 69.2%), or overall (71.4% vs. 69.5%) test scores compared to the control group ($n = 31$) based on the written test results. The authors determined that only 29% of physicians had systematic approaches to the interpretation of chest CTs.

Strengths of the Study. Although this was a negative study, this paper scored well in this critical appraisal because of the methodology. This was a multisite study, making the results more generalizable. A power calculation was performed to determine sample size. The participants were randomized into either a control or intervention group.

Relevance for Future Educational Advances. The lack of improvement in the ability to read chest CTs may be due to the ineffectiveness of a 1-hour educational intervention in transmitting knowledge to learners. Similar follow-up randomized studies are planned with a more comprehensive and structured educational approach to hopefully provide a more effective intervention.

Kessler CS, Afshar Y, Sardar G, Yudkowsky R, Ankel F, Schwartz A. A prospective, randomized, controlled study demonstrating a novel, effective model of transfer of care between physicians: the 5 Cs of consultation. Acad Emerg Med 2012;19:968–74.[36]
Background. Communication skills are important in EM and can mitigate medical error. The authors hypothesized that training using a standardized consultation communication protocol would improve communications and that the effect would be strongest for junior residents.

Methods. This was a single-blinded, single-institution, prospective, controlled trial of EM and EM/internal medicine residents who were randomized into either an intervention group that was trained in the consult protocol or an untrained group. Each participant placed recorded phone calls to a standardized consultant on two simulated patient cases. The recordings were rated by three blinded reviewers using a checklist instrument and by another two seasoned clinicians who used a global assessment scale. A power calculation suggested 17 residents per group to achieve significance.

Results. Forty-three of 47 eligible residents (91%) participated. There was excellent interrater reliability, and the intervention group performed significantly better than the control group. There was no diminution of effect by postgraduate year (PGY) level, suggesting that upper-level residents had not previously obtained the skills taught in the consultation protocol.

Strengths of the Study. This study has a rigorous methodology that can serve as a model for measuring the effect of an educational intervention under realistic conditions. The consultation protocol training was superior to no intervention.

Relevance for Future Educational Advances. Milestone assessment will depend on testing methods that can examine a specific skill set. Although cumbersome, this study presents a seemingly valid model.

Lee MO, Brown LL, Bender J, Machan JT, Overly FL. A medical simulation-based educational intervention for emergency medicine residents in neonatal resuscitation. Acad Emerg Med 2012;19:577–85.[41]
Background. EM residents have relatively little exposure to critically ill neonates. The objective of this study was to determine if a simulation-based educational intervention is a more effective teaching method than a standard didactic curriculum for neonatal resuscitation.

Methods. This single-center, randomized controlled study assessed the neonatal resuscitation knowledge and skills of EM residents. The study intervention was a 4-hour, simulation-based educational intervention that included didactics and several high-fidelity simulation scenarios, followed by expert video debriefing and procedural skills stations. A baseline and postintervention assessment was performed using 1) a questionnaire to evaluate confidence in leading adult, pediatric, and neonatal resuscitation and prior neonatal resuscitation experience and 2) a neonatal resuscitation simulation scenario to evaluate knowledge and skills. The control group received the standard EM curriculum. Assessments were recorded and reviewed independently by two evaluators using a validated neonatal resuscitation scoring tool. Outcomes measured included changes in overall neonatal resuscitation score, number of completed critical actions, time to initial steps of neonatal resuscitation, and changes in confidence level in leading a neonatal resuscitation.

Results. Twenty-seven of 36 residents were enrolled (12 intervention, 15 control). At the final assessment, the intervention group's neonatal resuscitation test score improved by a mean of 11.8% (p = 0.016) while the control group's score changed by –0.5% (p = 0.943). The intervention group performed 2.31 more critical actions, and the times to critical actions were also improved compared to controls. Furthermore, the proportion of residents who were "not at all confident" leading neonatal resuscitation decreased to 35% in the intervention group compared to 67% in the control group.

Strengths of the Study. This was a well-designed, randomized controlled study of a simulation-based intervention that used a previously developed measurement tool to assess knowledge and skill performance. The use of established assessment methods adds to the validity of the study rather than using a home-grown tool.

Relevance for Future Educational Advances. While there is still some debate about the role of simulation, this study joins a growing body of literature that suggests simulation-based education plays a role in teaching neonatal resuscitations, which are relatively rare in clinical practice.

Li CH , Kuan WS, Mahadevan M, Daniel-Underwood L, Chiu TF, Nguyen HB; ATLAS Investigators (Asia neTwork to reguLAte Sepsis care). A multinational randomised study comparing didactic lectures with case scenario in a severe sepsis medical simulation course. Emerg Med J 2012; 29:559–64.[43]
Background. Education in early goal-directed therapy is one of the barriers to implementation. This study seeks to understand the role of didactic lectures and simulation in training residents on the early management of severe sepsis.

Methods. This was a prospective, four-center, randomized study of sepsis education for EM residents in Asia using a crossover design. A 5-hour course was developed involving lectures and a skills and simulation workshop. Residents were block-randomized to lecture first or simulation first. Trainees were tested at three intervals. A pretest was given to the participants at the beginning of the course, posttest 1 was given after the didactic lectures or workshop/simulated case scenario depending on the study group assignment, and then a final posttest 2 was given at the end of the course after completing both the lectures and workshop. Performance on the simulated case scenario was evaluated with a performance task checklist.

Results. Ninety-eight participants were enrolled in the study. Pretest, posttest 1, and posttest 2 scores significantly improved in all participants (65.4%, SD ± 12.2%; 75.4%, SD ± 12.1%; and 80.8%, SD ± 12.0%, respectively; p < 0.01). Although there was no significant difference in posttest 2 scores between the two groups, the lecture-first group had significantly higher posttest 1 scores than the simulation-first group (78.8%, SD ± 10.6% vs. 71.6%, SD ± 12.6%; p < 0.01). Also, the final simulated case task performance completion rate was better in the lecture-first group (90.8% [SD ± 4.2%] vs. 83.8% [SD ± 4.3%], p = 0.02). These data support the fact that resident education in early goal-directed therapy should include a comprehensive curriculum that starts with didactic lectures followed by a simulation experience.

Strengths of the Study. This was a well-designed, multicenter, randomized controlled study using a block randomization and crossover design with an independent objective measure (written test and task performance completion) at each step of intervention. The crossover design allowed for a fair comparison of the different educational methodologies.

Relevance for Future Educational Advances. This study demonstrates that while simulation is an important teaching modality, it needs to be combined with traditional didactic education to optimize learning. It would have been helpful to calculate the effect size between the two groups to understand how much the learners improved.

Love JN, Howell JM, Hegarty CB, et al. Factors that influence medical student selection of an emergency medicine residency program: implications for training programs. Acad Emerg Med 2012;19:455–60.[46]
Background. Emergency medicine program directors desire an understanding of medical students' decision-making when choosing an EM residency program. The objective of this survey study was to identify and prioritize factors that medical students report influence their residency selection decisions.

Methods. This cross-sectional, multi-institutional study used an electronically delivered survey to anonymously collect responses from allopathic medical students over the 3-week period between the National Residency Matching Program rank list submission deadline and the announcement of match results. Sur-

vey questions were developed after a review of the pertinent literature, author discussion, and focus group information collected from incoming interns at three of the participating sites. Questions were pilot-tested with interns prior to being finalized. Authors rated the survey questions based on to what degree issues could be controlled by program directors, and the survey was distributed on a regional basis.

Results. Electronic survey invitations were sent to 1,525 students with a response rate of 57%. Ninety-six percent of respondents indicated that both geographic location (desire to be close to a partner or family) and independent program-specific attributes (interview experience and academic reputation) were important in residency choice. The authors noted that program-specific factors may be under the influence of program directors in enhancing a program's desirability.

Strengths of the Study. Although this study was of a survey design with a response rate less than 75%, it was included in this appraisal due to the rigor applied in the creation of the survey itself by the seven participating sites. Author discussion, review of the relevant literature, and focus groups held with incoming interns from three of the participating sites were the methods used to build validity into the survey instrument.

Relevance for Future Educational Advances. Well-constructed surveys can be effective in education research when careful attention is paid to the validation of survey content and response processes, as well as thoughtful survey distribution.

Pusic MV, Andrews JS, Kessler DO, et al. Prevalence of abnormal cases in an image bank affects the learning of radiograph interpretation. Med Educ 2012;46:289–98.[55]
Background. The authors created an online teaching module for reviewing pediatric ankle radiographs. They hypothesized that the ratio of normal to abnormal training cases would affect learner outcomes.

Methods. This was a multi-institutional, prospective, double-blind, randomized, three-arm trial using pediatric, EM, and pediatric EM residents. The authors developed an online, self-administered, untimed radiology training set, where residents were to identify each of 50 pediatric ankle images as normal or abnormal, as well as mark the injury location. On each case, the learner received immediate feedback consisting of the radiologist report and highlighting of any injury. The residents were randomly divided into one of three versions of the training set, which contained 30, 50, or 70% abnormal cases. All participants then completed a 20-image posttest with an abnormal case prevalence of 40%, which was considered similar to that seen in clinical practice. A power analysis suggested a minimum requirement of 90 participants.

Results. One-hundred of 355 (28%) eligible residents from six institutions participated; two-thirds of them were pediatric residents. The accuracy of performance on the posttests was similar for each group and showed a significant improvement from the training-set performance; however, the groups showed a significantly different sensitivity–specificity trade-off in achieving that accuracy. The group with 70% abnormal images in its training set had a higher sensitivity and lower specificity and therefore higher false-positive rate compared to the groups with 50 and 30% abnormal images. In contrast, the 30% group was more specific and less sensitive than the other groups resulting in the same accuracy, but a higher false-negative rate.

Strengths of the Study. This was a well-designed, prospective, multi-institutional, multispecialty study with an objective endpoint, whose enrollment exceeded the target sample size, and the results demonstrated statistical significance.

Relevance for Future Educational Advances. In addition to being a role model for technically superior methodology in testing an educational intervention, the study demonstrates the care that needs to be used in selecting a clinically relevant balance of normal to abnormal pathology in a training module.

Bernard AW , Kman NE,Way DP , Khandelwal S. The impact of two clinical shift allocation models on student experiences in an emergency medicine clerkship. Teach Learn Med 2012;24:194–9.[19]
Background. Medical students' clinical schedules during an EM clerkship have traditionally consisted of a mixture of day, evening, and night shifts independent of EM faculty schedules, which often leads to teacher–learner discontinuity in the ED. The study compared the traditional model of student shift scheduling with a "continuity-based shift model" intended to maximize students' clinical time with only one to three faculty members.

Methods. This qualitative, prospective, crossover cohort study was conducted in one institution over 4 months. All students completed 2 weeks under the "traditional shift model" and 2 weeks under the "continuity-based shift model." At the end of each 2-week block, students completed a 10-item survey about their experience and participated in a semistructured group interview. Survey responses were matched and analyzed using two-way analysis of variance. Interviews were recorded, transcribed, and analyzed for emerging themes using appropriate qualitative methods.

Results. Students ($n = 18$) rated the continuity shift model more highly, regardless of the order in which they participated in the 2-week blocks. Ratings of faculty teaching, interaction, and feedback were significantly higher during the continuity shift model. Six themes and 16 subthemes emerged from the analysis of the group interviews. Two unique themes of feedback and the teacher–learner relationship were superior in the continuity shift model, although teaching was noted to be very attending-dependent regardless of either model.

Strengths of the Study. This mixed-methods study triangulated quantitative survey data and qualitative data collection and analysis. Careful attention was paid to the qualitative data collection and analysis phase, including refinement of emerging themes and independent theme validation by up to four investigators.

Relevance for Future Educational Advances. Maximizing students' learning in the ED may require novel approaches in scheduling, such that temporal and faculty continuity exposures are balanced. Mixed-methods research can enhance our understanding of students' learning experiences in this complex, multifaceted environment.

Lin CS, Chiu TF, Yen DHT, Chong CF. Mini-clinical evaluation exercise and feedback on postgraduate trainees in the emergency department: a qualitative content analysis. Acute Med 2012;2:1–7.[45]

Background. The study examined the quality of feedback provided to first-year residents (PGY-1) by preceptors during mini–clinical evaluation exercises (mini-CEXs) in the ED.

Methods. This prospective, observational cohort study included 20 teaching hospitals in Taiwan. Residents were observed during mini-CEXs performing focused histories and physical examinations under the direct observation of board-certified emergency physician preceptors, who had completed a faculty-training course on clinical teaching, evaluation, and feedback. In addition to scoring these residents in seven domains, the preceptors provided positive and negative feedback and an action plan for each resident. Feedback in the evaluation forms was analyzed qualitatively using the constant comparative method based on grounded theory and coded into unique domains. Sampling continued until novel information was no longer generated.

Results. Mini-CEX sessions ($n = 983$) were collected on 230 PGY-1 residents from 242 preceptors with 85.3% of sessions providing feedback. Items from the evaluation forms were grouped into seven domains of clinical competence. The domain receiving the most comments (50.4%) from preceptors was "clinical judgment." Areas receiving the least amount of feedback were in the areas of communication and professionalism.

Strengths of the Study. This was a very large multicenter study that examined the frequency, type, and quality of feedback provided by preceptors. In this qualitative study, careful attention was paid to the data analysis phase, which included maintaining a careful audit trail and obtaining data saturation, which are both necessary for qualitative research validation strategies.

Relevance for Future Educational Advances. The mini-CEX can be used in the ED setting with the AC-GME competencies serving as a structure for feedback. The tool may need further modification to ensure preceptors facilitate the development of all needed skills, including reflection and communication. Analysis of a structured evaluation tool can provide valuable feedback to faculty in ongoing faculty development to improve evaluation of learners.

Trends in Medical Education Research in 2012

This year marks the fifth year of our review of medical education research that focuses on topics of interest to EM educators. For the current year, 48 publications met our review criteria. Fourteen (29%) received funding, including five of the nine highlighted articles (55.5%). The sources of funding were as follows: five federal (two NIH, one Canada, and one Taiwan), five university-sponsored, two industry-supported, and two organizationally funded. Research methodology included 17 surveys (36%),[17,19,21,25,30,31,33,37,40,42,46,50–52,54,60,61] 15 (31%) observational analyses,[14–16,22,23,27,32,33,38,44,47,49,56,57,59] and three (6%) qualitative methodology studies.[19,28,45] There were only 12 (25%) with an experimental or quasi-experimental study design,[20,24,26,35,36,39,41,43,48,53,55,57] with five of the highlighted articles using this rigorous design.[35,36,41,43,55] These trends are summarized in Table 3.

Residents were the subjects for 38 (79%) of all studies and seven (78%) of those highlighted here. In some cases, resident learners were combined with medical students or faculty subjects. Publication in EM journals predominated (36; 75%). Three (6%) appeared in medical education journals, and nine (19%) were published

Table 3
Trends for the Reviewed Education Research Articles of 2012

Variable	All Publications ($n = 48$)	Highlighted ($n = 9$)
Funding	14	5
Learner group*		
Medical students	11	2
Residents	38	7
Other	13	1
Study methodology		
Survey	17	1
Observational	15	1
Experimental/quasi-experimental	13	5
Qualitative	3	2
Prevalent topics of study*		
Learner satisfaction/confidence	31	5
Technology	22	4
Simulation	16	3
Competency of learners	15	3
Pediatrics	9	2
Learning methods	8	2
Procedures for EM	7	0
Communication	5	1
Ultrasound	5	0
Disaster	5	0
Psychobehavioral	5	0
Location of study		
United States	26	4
Canada	8	2
Other	14	3

*It is possible to exceed 100% for these categories because of multiple populations or study topics.

in journals ranging from surgery[33,42,44,54] to psychiatry.[31,52] Others included palliative medicine,[26] critical care,[56] and simulation.[38] Most articles (81%) had authors with an EM affiliation. Five (10%) partnered with a specialist in medical education and 15 (31%) collaborated with a faculty member from another discipline.

Technology, especially simulation, maintained its prevalence in 22 (46%) of the medical education studies in 2012. Simulation accounted for 16 (33%) of the studies and three of those were highlighted.[33,41,43] Seven of these focused on learner satisfaction or self-assessment regarding the simulation experience, while nine were outcomes based.[22,32,33,38,39,41,43,56,57] Five articles studied ultrasound in EM,[21,47,49,59,61] and two employed visual technology.[55,58]

A positive trend this year was the emergence of more competency-based studies, which reflect a higher order of research outcomes in medical education, based on the Kirkpatrick model.[3,4] Fifteen (31%) of the articles reviewed, including three of the highlighted articles, had objectively measurable outcomes.[33,45,55] Seven of these 15 studies received funding. Studies measuring learner satisfaction, self-assessment, or comfort level (lowest Kirkpatrick tier) were featured in 31 (65%) of the 2012 articles. Five of these articles appeared in our highlighted section, but for each one of those featured, this metric was a secondary outcome measure.[18,33,41,45,46]

Pediatric topics accounted for nine (19%) of the articles reviewed.[20–22,24,40,41,55,56,58] The two that are in the highlighted section used technology as the basis of their studies, and both received funding.[41,55] Each of the articles with a pediatric focus had the primary or supporting author with specialized pediatric training. Other prevalent subject areas with specialty author collaboration were psychiatry,[23,31,50,52,54] education,[15,16,18,20,33,34,36,57] and disaster medicine,[22,30,34,36,42] with all but one using simulation in their studies. Five articles (10%) this year evaluated tools that were developed for educational purposes[14,29,32,36,56] and learning methods were evaluated in eight articles (17%).[19,21,30,35,38,43,44,49]

DISCUSSION

Upon reflection of the past 5 years of reviewing medical education research, the authors have noted several interesting features. Of note, the prevalence of funded medical education studies has diminished somewhat (Figure 1). Funded studies are prominent in the articles highlighted for superior methodology. This is consistent with findings by Reed et al.[5] who noted that funded studies were more likely to have been of higher quality as assessed on a validated scale. The trend in study methodologies favored observational and survey design (Figure 2). When we attempted to observe trends in topic areas of study, we noted that while there may have been a theme in a given year, there is not a trend in any area other than an increase in the use of technology, particularly simulation, to address a variety of topics (Figure 3). Most often, simulations focus on clinical issues related to critical patients that are rarely available to learners. Additionally, simulations are being used in the assessment of learners.
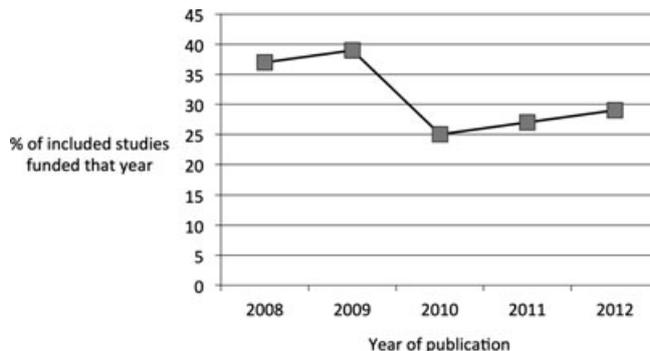
Figure 1. Percentage of EM education research publications that were funded (2008–2012).
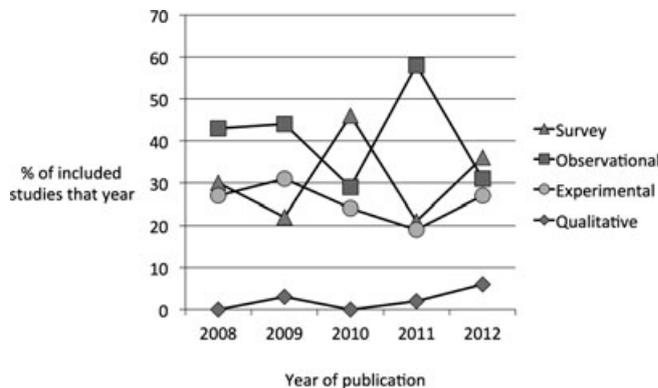
Figure 2. Types of study designs employed in published EM education research (2008–2012).
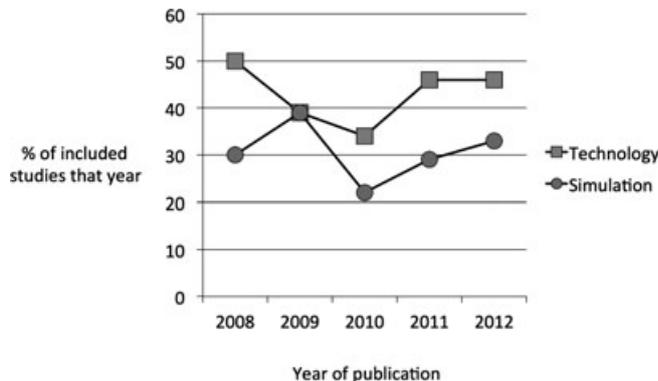
Figure 3. Percentage of EM education research publications that involved technology and specifically simulation (2008–2012).

Qualitative Studies

This year, the authors recognized the small but growing number of educational research studies employing qualitative methods of data collection and analysis. Although rather new to the field of EM education research, qualitative research can contribute important theory-driven knowledge that explores, informs, and expands our understanding of how and why existing practices work. When performed well, qualitative studies can further our understanding of complex processes such as learn-

ing and can uncover new areas for advanced study. Qualitative research studies can be performed and assessed using standards that are parallel to those of quantitative research, including the demands for ethical considerations, rigorous methodology, credibility, and relevance.

A second scoring system was thus warranted to appropriately assess qualitative studies. The inherent challenges of judging the quality of qualitative studies prompted a search for literature-based metrics on which to base this scoring system. Using accepted recommendations and hierarchical formulations, a qualitative scoring system was derived, which allowed for the appraisal of papers based on both theoretical and technical grounds and the central methodological procedures of qualitative research.[10–13]

LIMITATIONS

A limitation in this critical appraisal series includes the fact that the names of authors from the included articles were not blinded from the critical appraisal reviewer team, which may have biased the scores. This limitation was minimized by excluding the reviewer from scoring publications where there was potentially a conflict of interest (own publication, own institution, or had a vested interest in the authors or work).

Additionally, two changes were made to the methodology in this year's critical appraisal of the literature, which may have resulted in two additional limitations beyond those cited in previous years. First, single-site surveys, which measured satisfaction or self-assessment using unvalidated instruments, were removed from the final publication list to be formally reviewed by the entire six-person panel, because these reaction-based studies are the least rigorous per the Kirkpatrick model. This may have resulted in our erroneously omitting high-quality studies, although no single-site, survey-based studies published in 2008 through 2011 were highlighted as superior studies in this series over the past 4 years.[6–9] Second, a new scoring system was used to critically appraise qualitative studies. Although this has not been validated, it closely mirrors the same domains as the scoring system for quantitative studies and was derived using accepted measures based on the literature.[10–13] The unique metrics included attention to theoretical underpinning of study design, auditing of collected data, techniques of analysis (including triangulation and validation of emerging themes), and the relationship between a study's author and the research participants. Although particular to qualitative research, these metrics parallel the rigorous standards for validity, reliability, and attention to bias commonly required in quantitative research.

CONCLUSIONS

This critical appraisal of the EM education research literature highlights quality publications and recent trends in the field. The seven quantitative and two qualitative studies featured represent methodologically superior research published in 2012. Each contributes to the expanding field of education research, while addressing the methods to control, justify, or minimize the limitations that are inherent to this focus. These highlighted studies can serve as exemplary models for emergency medicine educators interested in conducting high-quality, methodologically sound education research.

References

1. LaMantia J, Deiorio NM, Yarris LM. Executive summary: education research in emergency medicine-opportunities, challenges, and strategies for success. Acad Emerg Med 2012;19:1319–22.
2. Boyer E. Scholarship Reconsidered: Priorities of the Professoriate. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching, 1990.
3. Hutchinson L. Evaluating and researching the effectiveness of educational interventions. Br Med J 1999;318:1267–9.
4. Kirkpatrick DL. Evaluation of training. In: Craig RL, ed. Training and Development Handbook: A Guide to Human Resource Development. New York, NY: McGraw Hill, 1976.
5. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA 2007;298:1002–9.
6. Farrell SE, Coates WC, Khun GJ, Fisher J, Shayne P, Lin M. Highlights in emergency medicine medical education research: 2008. Acad Emerg Med 2009;16:1318–24.
7. Kuhn GJ, Shayne P, Coates WC, et al. Critical appraisal of emergency medicine educational research: the best publications of 2009. Acad Emerg Med 2010;17(Suppl 2):S16–25.
8. Shayne P, Coates WC, Farrell SE, et al. Critical appraisal of emergency medicine educational research: the best publications of 2010. Acad Emerg Med 2011;18:1081–9.
9. Fisher J, Lin M, Coates WC, et al. Critical appraisal of emergency medicine educational research: the best publications of 2011. Acad Emerg Med 2013;20:200–8.
10. Coté L, Turgeon J. Appraising qualitative research articles in medicine and medical education. Med Teach 2005;27:71–5.
11. Daly J, Willis K, Small R, et al. A hierarchy of evidence for assessing qualitative health research. J Clin Epidemiol 2007;60:43–9.
12. Jeanfreau SG, Jack L Jr. Appraising qualitative research in health education: guidelines for public health educators. Health Promot Pract 2010;11:612–7.
13. Dixon-Woods M, Sutton A, Shaw R, et al. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. J Health Serv Res Policy 2007;12:42–7.
14. Ander DS, Wallenstein J, Abramson JL, Click L, Shayne P. Reporter-Interpreter-Manager-Educator (RIME) descriptive ratings as an evaluation tool in an emergency medicine clerkship. J Emerg Med 2012;43:720–7.
15. Aram N, Brazil V, Davin L, Greenslade J. Intern underperformance is detected more frequently in

emergency medicine rotations. Emerg Med Australas 2013;25:68–74.

16. Atzema CL, Stefan RA, Saskin R, Michlik G, Austin PC. Does ED crowding decrease the number of procedures a physician in training performs? A prospective observational study. Am J Emerg Med 2012;30:1743–8.

17. Avegno JL, Murphy-Lavoie H, Lofaso DP, Moreno-Walton L. Medical students' perceptions of an emergency medicine clerkship: an analysis of self-assessment surveys. Int J Emerg Med 2012;5:25.

18. Bernard AW, Kman NE, Betz B, Khandelwal S, Caterino JM. Faculty prefer continuity with medical students in the emergency department. Emerg Med J 2013;30:327–8.

19. Bernard AW, Kman NE, Way DP, Khandelwal S. The impact of two clinical shift allocation models on student experiences in an emergency medicine clerkship. Teach Learn Med 2012;24:194–9.

20. Bhanji F, Gottesman R, de Grave W, Steinert Y, Winer LR. The retrospective pre-post: a practical method to evaluate learning from an educational program. Acad Emerg Med 2012;19:189–94.

21. Bretholz A, Doan Q, Cheng A, Lauder G. A presurvey and postsurvey of a web- and simulation-based course of ultrasound-guided nerve blocks for pediatric emergency medicine. Pediatr Emerg Care 2012;28:506–9.

22. Cicero MX, Auerbach MA, Zigmont J, Riera A, Ching K, Baum CR. Simulation training with structured debriefing improves residents' pediatric disaster triage performance. Prehosp Disaster Med 2012;27:239–44.

23. Cinar O, Ak M, Sutcigil L, et al. Communication skills training for emergency medicine residents. Eur J Emerg Med 2012;19:9–13.

24. Corwin DJ, Kessler DO, Auerbach M, Liang A, Kristinsson G. An intervention to improve pain management in the pediatric emergency department. Pediatr Emerg Care 2012;28:524–8.

25. Damon Dagnone J, Takhar A, Lacroix L. The Simulation Olympics: a resuscitation-based simulation competition as an educational intervention. CJEM 2012;14:363–8.

26. DeVader TE, Jeanmonod R. The effect of education in hospice and palliative care on emergency medicine residents' knowledge and referral patterns. J Palliat Med 2012;15:510–5.

27. Fix ML, Enslow MS, Blankenship JF, et al. Emergency medicine resident anesthesia training in a private vs. academic setting. J Emerg Med 2013;44:676–81.

28. Flowerdew L, Brown R, Russ S, Vincent C, Woloshynowych M. Teams under pressure in the emergency department: an interview study. Emerg Med J 2012;29:e2.

29. Flowerdew L, Brown R, Vincent C, Woloshynowych M. Development and validation of a tool to assess emergency physicians' nontechnical skills. Ann Emerg Med 2012;59:376–85.

30. Franc JM, Nichols D, Dong SL. Increasing emergency medicine residents' confidence in disaster management: use of an emergency department simulator and an expedited curriculum. Prehosp Disaster Med 2012;27:31–5.

31. Gordon JT. Emergency department junior medical staff's knowledge, skills and confidence with psychiatric patients: a survey. Psychiatrist 2012;36:186–8.

32. Hall AK, Pickett W, Dagnone JD. Development and evaluation of a simulation-based resuscitation scenario assessment tool for emergency medicine residents. CJEM 2012;14:139–46.

33. Harvey A, Bandiera G, Nathens AB, LeBlanc VR. Impact of stress on resident performance in simulated trauma scenarios. J Trauma Acute Care Surg 2012;72:497–503.

34. Hicks CM, Kiss A, Bandiera GW, Denny CJ. Crisis resources for emergency workers (CREW II): results of a pilot study and simulation-based crisis resource management course for emergency medicine residents. CJEM 2012;14:354–62.

35. Keijzers G, Sithirasenan V. The effect of a chest imaging lecture on emergency department doctors' ability to interpret chest CT images: a randomized study. Eur J Emerg Med 2012;19:40–5.

36. Kessler CS, Afshar Y, Sardar G, Yudkowsky R, Ankel F, Schwartz A. A prospective, randomized, controlled study demonstrating a novel, effective model of transfer of care between physicians: the 5 Cs of consultation. Acad Emerg Med 2012;19:968–74.

37. Kim TE, Reibling ET, Denmark KT. Student perception of high fidelity medical simulation for an international trauma life support course. Prehosp Disaster Med 2012;27:27–30.

38. Kobayashi L, Dunbar-Viveiros JA, Devine J, et al. Pilot-phase findings from high-fidelity in situ medical simulation investigation of emergency department procedural sedation. Simul Healthc 2012;7:81–94.

39. Kyaw Tun J, Granados A, Mavroveli S, et al. Simulating various levels of clinical challenge in the assessment of clinical procedure competence. Ann Emerg Med 2012;60:112–20.

40. Lee JA, Chernick L, Sawaya R, Roskind CG, Pusic M. Evaluating cost awareness education in US pediatric emergency medicine fellowships. Pediatr Emerg Care 2012;28:655–75.

41. Lee MO, Brown LL, Bender J, Machan JT, Overly FL. A medical simulation-based educational intervention for emergency medicine residents in neonatal resuscitation. Acad Emerg Med 2012;19:577–85.

42. Leow JJ, Brundage SI, Kushner AL, et al. Mass casualty incident training in a resource-limited environment. Br J Surg 2012;99:356–61.

43. Li CH, Kuan WS, Mahadevan M, Daniel-Underwood L, Chiu TF. Nguyen HB; ATLAS Investigators (Asia neTwork to reguLAte Sepsis care). A multinational randomised study comparing didactic lectures with case scenario in a severe sepsis medical simulation course. Emerg Med J 2012;29:559–64.

44. Lifchez SD. Hand education for emergency medicine residents: results of a pilot program. J Hand Surg Am 2012;37:1245–8.

45. Lin CS, Chiu TF, Yen DH, Chong CF. Mini-clinical evaluation exercise and feedback on postgraduate

trainees in the emergency department: a qualitative content analysis. J Acute Med 2012;2:1–7.

46. Love JN, Howell JM, Hegarty CB, et al. Factors that influence medical student selection of an emergency medicine residency program: implications for training programs. Acad Emerg Med 2012; 19:455–60.

47. MacVane CZ, Irish CB, Strout TD, Owens WB. Implementation of transvaginal ultrasound in an emergency department residency program: an analysis of resident interpretation. J Emerg Med 2012;43:124–8.

48. Mahler SA, McCartney JR, Swoboda TK, Yorek L, Arnold TC. The impact of emergency department overcrowding on resident education. J Emerg Med 2012;42:69–73.

49. Mahler SA, Swoboda TK, Wang H, Arnold TC. Dedicated emergency department ultrasound rotation improves residents' ultrasound knowledge and interpretation skills. J Emerg Med 2012;43:129–33.

50. Marciano R, Mullis DM, Jauch EC, et al. Does targeted education of emergency physicians improve their comfort level in treating psychiatric patients? West J Emerg Med 2012;13:453–7.

51. Marco CA, Kowalenko T. Competence and challenges of emergency medicine training as reported by emergency medicine residents. J Emerg Med 2012;43:1103–9.

52. Morrison A, Roman B, Borges N. Psychiatry and emergency medicine: medical student and physician attitudes toward homeless persons. Acad Psychiatry 2012;36:211–5.

53. Na JU, Sim MS, Jo IJ, Song HG, Song KJ. Basic life support skill retention of medical interns and the effect of clinical experience of cardiopulmonary resuscitation. Emerg Med J 2012;29:833–7.

54. Peckler B, Prewett MS, Campbell T, Brannick M. Teamwork in the trauma room evaluation of a multimodal team training program. J Emerg Trauma Shock 2012;5:23–7.

55. Pusic MV, Andrews JS, Kessler DO, et al. Prevalence of abnormal cases in an image bank affects the learning of radiograph interpretation. Med Educ 2012;46:289–98.

56. Reid J, Stone K, Brown J, et al. The Simulation Team Assessment Tool (STAT): development, reliability and validation. Resuscitation 2012;83:879–86.

57. Smith S, Tallentire V, Wood M, Cameron H. The distracted intravenous access (DIVA) test. Clin Teach 2012;9:320–4.

58. Srivastava G, Roddy M, Langsam D, Agrawal D. An educational video improves technique in performance of pediatric lumbar punctures. Pediatr Emerg Care 2012;28:12–6.

59. Torres-Macho J, Antón-Santos JM, García-Gutierrez I, et al. Working Group of Clinical Ultrasound, Spanish Society of Internal Medicine. Initial accuracy of bedside ultrasound performed by emergency physicians for multiple indications after a short training period. Am J Emerg Med 2012;30:1943–9.

60. Wittels KA, Takayesu JK, Nadel ES. A two-year experience of an integrated simulation residency curriculum. J Emerg Med 2012;43:134–8.

61. Zaia BE, Briese B, Williams SR, Gharahbaghian L. Use of cadaver models in point-of-care emergency ultrasound education for diagnostic applications. J Emerg Med 2012;43:683–91.