A Randomized Comparison Trial of Case-based Learning versus Human Patient Simulation in Medical Student Education

Lawrence R. Schwartz, MD, Rosemarie Fernandez, MD, Sarkis R. Kouyoumjian, MD, Kerin A. Jones, MD, Scott Compton, PhD

Abstract

Objectives: Human patient simulation (HPS), utilizing computerized, physiologically responding mannequins, has become the latest innovation in medical education. However, no substantive outcome data exist validating the advantage of HPS. The objective of this study was to evaluate the efficacy of simulation training as compared with case-based learning (CBL) among fourth-year medical students as measured by observable behavioral actions.

Methods: A chest pain curriculum was presented during a one-month mandatory emergency medicine clerkship in 2005. Each month, students were randomized to participate in either the CBL-based or the HPS-based module. All students participated in the same end-of-clerkship chest pain objective structured clinical examination that measured 43 behaviors. Three subscales were computed: history taking, acute coronary syndrome evaluation and management, and cardiac arrest management. Mean total and subscale scores were compared across groups using a multivariate analysis of variance, with significance assessed from Hotelling’s $T^2$ statistic.

Results: Students were randomly assigned to CBL ($n = 52$) or HPS ($n = 50$) groups. The groups were well balanced after random assignment, with no differences in mean age (26.7 years; range, 22–44 years), gender (male, 52.0%), or emergency medicine preference for specialty training (28.4%). Self-ratings of learning styles were similar overall: 54.9% were visual learners, 7.8% auditory learners, and 37.3% kinetic learners. Results of the multivariate analysis of variance indicated no significant effect (Hotelling’s $T^2 [3,98] = 0.053; p = 0.164$) of education modality (CBL or HPS) on any subscale or total score difference in performance.

Conclusions: HPS training offers no advantage to CBL as measured by medical student performance on a chest pain objective structured clinical examination.

Keywords: medical education, simulation, randomized-comparison trial

Over the past decade, the use of simulation in medical education has increased exponentially. The term “simulation” spans a wide variety of formats, from the low-tech actor portraying a standardized patient to high-fidelity mannequin-based human patient simulation (HPS). HPS is capable of both simulating realistic patient encounters and giving real-time, physiologically accurate feedback. With simulators mirroring human disease states, students are able to practice care of critically ill patients in a realistic setting without risk of harm to actual patients. HPS has the added advantage of being available whenever needed and does not rely on random patient encounters for medical education. As Gordon and Pawlowski aptly described, simulation can provide “education on demand.”1 The appeal of such a safety-oriented, efficient teaching modality is clear and explains the relatively rapid incorporation of this technology into medical training curricula.

The general acceptance of simulation as a beneficial training modality is reflected in the current literature. Overwhelmingly, student participants state that they enjoy simulation-based training and believe that it prepares them well for clinical practice.2,3 Medical educators comment on the realism and accuracy with which simulation replicates clinical experience.4 While participants respond favorably to HPS-based training, there is a paucity of evidence supporting the superiority of simulation versus more traditional teaching formats.5,6 Absence of
such evidence limits the ability of educators to justify the upfront costs associated with investing in this technology. In 2005, Issenberg et al. published a systematic literature review defining several uses of HPS that lead to effective learning. While such support for this technology is impressive, the overall characteristics of available research were not. Most studies included in the review were nonrandomized, and most contained fewer than 30 total participants. Not surprisingly, most articles included were published in specialty-specific journals. These study flaws are not limited to simulation literature and have been described in the general medical education research. Recently, the medical community has called for more stringent standards in medical education research. The simulation community as a whole has recognized the need to apply outcomes-based research to the field of simulation.

The goal of this study was to evaluate the efficacy of simulation training as compared with standard case-based learning (CBL) among fourth-year medical students as measured by observable behavioral actions. A randomized study comparing CBL teaching with HPS-based training of acute coronary syndrome (ACS) management was performed. Performance on a chest pain objective structured clinical examination (OSCE) was used as the outcome measure.

METHODS

Study Design
This was a prospective, randomized study comparing performance on an ACS OSCE between students taught with CBL classroom instruction and HPS. Prior expedited approval was obtained from the Wayne State University Human Investigation Committee. All participants gave written informed consent.

Study Setting and Population
This study was conducted at Wayne State University School of Medicine (WSU-SOM) in Detroit, Michigan. The WSU-SOM is the third largest medical school in the country and the largest single campus medical school, with a total student body of more than 1,000 students. The simulation center is located within the Eugene Applebaum College of Pharmacy and Health Sciences, which occupies a 4,000-square foot area consisting of four HPS rooms and other peripheral rooms (locker room, three debriefing rooms). All simulation training was performed using Medical Education Technologies, Inc. (Sarasota, FL) HPS.

From July 2005 to November 2005, a total of 105 fourth-year medical students enrolled in a mandatory, month-long emergency medicine (EM) clerkship. All participants had completed the same third year curriculum and were enrolled in the first half of the fourth year. In the first week of the clerkship, all students were given a lecture on EM management of acute chest pain and were given the core objectives (Table 1) and reading material, which covered the management of ACS. The same individual (LRS) gave the lecture throughout the year. Students who failed to attend the lecture were excluded from the study.

In the second week of the clerkship, students gave consent for participation in the study and were randomized to receive either HPS-based instruction or a CBL session that had previously been part of the curriculum. Those students who did not consent or who were excluded were still included in the randomization. Additionally, all students agreed to abide by a standard confidentiality contract used by the Eugene Applebaum College of Pharmacy and Health Sciences Simulation Center. Failing to abide by the contract was considered a violation of the WSU-SOM Honor Code.

Study Protocol
At the start of the EM clerkship, each fourth-year student filled out a questionnaire designed to elicit general demographics (age, gender), resuscitation experience (witness or medical team participant), choice of residency specialty, and self-determined learning style. Learning style was determined using a tool based on the Barsch Learning Styles Inventory. Students were given a chart describing visual, auditory, and kinesthetic learning styles and asked to self-select which described them best.

CBL Session. During the CBL learning session, students were presented with a vignette describing a patient with classic signs of ACS. Student participation was solicited to determine the proper history taking, workup, management, and disposition of the patient. During this session, the facilitator (SRK) covered the cardiac chest pain objectives listed in the WSU-SOM EM medical student curriculum (Table 1). Following the discussion component, students reviewed the advanced cardiac life support protocols for ventricular tachycardia and ventricular fibrillation using a LifePak 9 Cardiac Monitor (Medtronic, Redmond, WA) and RhythmSIM cardiac rhythm generator (model AA-750; Armstrong Medical Industries, Inc., Lincolnshire, IL). The total time spent in the CBL session was approximately 1 hour.

HPS Session. Students randomized to receive HPS-based training were given a 15-minute orientation to the mannequin, during which the available equipment and specialized features of the simulation center were explained. The entire group was then divided into smaller groups of four to six students each. Each group was called in individually to assess, manage, and disposition a simulated patient with ACS and subsequent ventricular tachycardia or ventricular fibrillation arrest. All cardiac rhythms were projected from the HPS mannequin to a LifePak Cardiac Monitor identical to the one used during the CBL session. The simulated case was identical to the one used in the standard (CBL) small group discussion and was designed to teach and review the cardiac chest pain objectives listed in the curriculum (Table 1). An instructor (LRS) playing the role of a nurse was present in the room throughout the scenario to assist students with tasks and provide general guidance when needed. No formal instruction was given during the scenario. After all three groups performed the simulation, the entire group was convened for debriefing of the case. The size of the group debriefed mirrored that of the CBL group. The instructor-led discussion (RF) was designed to review correct management of the
case and point out positive team performance. The content of the debriefing was similar to that of the CBL session but was presented in the context of the performance of the groups. Student participation was actively solicited and self-critique encouraged. The total teaching time for the simulation session was approximately 1 hour.

**OSCE.** At the end of the clinical rotation, all students took part in an ACS OSCE. The scenario was similar to the case presented earlier in the curriculum. An actor portraying a patient with ACS allowed the students to take a focused history and plan their workup. The cardiac arrest component of the OSCE was staged using a Resusci Anne mannequin (model 31000501; Laerdal, Wappingers Falls, NY) and a LifePak Cardiac Monitor. A trained evaluator blinded to intervention groups scored the students’ performance on assessment, workup, and treatment of the standardized patient utilizing a 43-point checklist of required actions. To verify interrater reliability, the sessions were recorded on DVD and scored by physicians who were also blinded to the students’ training protocol.

**Data Analysis**
Descriptive statistics were generated by group regarding participant demographics (such as age and gender). Differences in the primary outcomes, which included the three subscale scores from the OSCE, were assessed using multivariate analysis of variance (MANOVA) to control for type I error inflation. Significance was assessed from Hotelling’s $T^2$ statistic. The OSCE examination was comprised of 43 behavioral observations, and the three subscales were history ($n = 22$), AMI evaluation and management ($n = 13$), and cardiac arrest management ($n = 8$). All statistical analyses were performed using SPSS version 14.0 (SPSS Inc., Chicago, IL). For the purpose of study planning, a moderate-to-large effect size (0.75 of a standard deviation) was posited as the minimally meaningful difference in achievement means between groups. Using the sample size formula proposed by Lauter for MANOVA study designs, it was determined that a minimum of 46 participants would be needed in each group to achieve strong power (0.80), with the significance level set at 0.05.11

**RESULTS**
Students at the WSU-SOM attending the mandatory fourth-year EM elective during the months of September, October, and November 2005 were included in this study. Only three students were excluded due to failure to attend the lecture component of the intervention. All
eligible students consented to participate (Figure 1). Each month, students were randomly assigned to the standard CBL (n = 52) or HPS (n = 50) groups. The groups were well balanced in respect to mean age, percent male, and specialty training preference. Due to sample size limitations, this study was unable to statistically control for many factors that could potentially impact the primary outcome. Some of these variables were recorded and compared between groups. There was no evidence to suggest that differences existed between groups with respect to the students’ specialty interest at the time of the study (p = 0.844), experience with cardiac arrest codes in hospital settings (p = 0.116), and also self-rated learning style (p = 0.558). However, six students in the HPS group had not witnessed a medical resuscitation attempt compared with none of the students in the standard group (p = 0.012). Table 2 shows participant characteristics by group randomization.

Results of the MANOVA indicated no significant effect (Hotelling’s $T^2 [3,98] = 0.053; p = 0.164$) of education modality (standard or HPS) on the difference in the mean number of observed actions performed on any of the three subscales. Additionally, there was no overall mean difference between groups (p = 0.770). As can be seen by the Figures, student performance was very similar between groups for a majority of the items. Table 3 describes the group performance on each scale. A visual depiction of the percentage of students who performed each of the desired actions within each of the subscales is provided in Figures 2–4.

A subset (n = 20) of the evaluators’ ratings was corroborated by physician raters (RF and SRK). The overall percent agreement between the physician and trained evaluator scores was 89%. Cronbach’s $\alpha$ is provided as a measure of internal consistency for each scale and is displayed in Table 4.

**DISCUSSION**

Our objective was to evaluate the efficacy of simulation-based training as compared with CBL. We performed a randomized comparison study of medical simulation and CBL sessions and found no significant difference in outcomes as measured by performance on a clinically relevant OSCE. We sought to design a study that would contribute quantitative data to the growing body of simulation literature and be applicable to a wide range of learners.

This study design had several strengths. First and foremost, it is one of only a few prospective randomized studies comparing simulation and other teaching modalities in the medical literature. Wayne et al. reported the results of a randomized trial showing that simulation-based training improved learner performance on an advanced cardiac life support scenario after simulator training.\(^\text{12}\) This study used a wait-list control group

---

**Table 2**

<table>
<thead>
<tr>
<th>Participant Characteristics</th>
<th>Case-based Learning (n = 52)</th>
<th>Human Patient Simulation (n = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male, n (%)</td>
<td>26 (50.0)</td>
<td>27 (54.0)</td>
</tr>
<tr>
<td>Mean age, yr (standard deviation)</td>
<td>26.9 (3.8)</td>
<td>26.5 (3.2)</td>
</tr>
<tr>
<td>Planned specialties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal medicine</td>
<td>4 (7.7)</td>
<td>8 (16.0)</td>
</tr>
<tr>
<td>Emergency medicine</td>
<td>15 (28.8)</td>
<td>14 (28.0)</td>
</tr>
<tr>
<td>Obstetrics/gynecology</td>
<td>4 (7.7)</td>
<td>2 (4.0)</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>5 (9.6)</td>
<td>2 (4.0)</td>
</tr>
<tr>
<td>Radiology</td>
<td>2 (3.8)</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>Anesthesiology</td>
<td>4 (7.7)</td>
<td>8 (16.0)</td>
</tr>
<tr>
<td>Surgery</td>
<td>7 (13.5)</td>
<td>2 (4.0)</td>
</tr>
<tr>
<td>Family medicine</td>
<td>5 (9.6)</td>
<td>2 (4.0)</td>
</tr>
<tr>
<td>Neurology</td>
<td>1 (1.9)</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>2 (3.8)</td>
<td>4 (8.0)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (5.8)</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>Experienced medical resuscitation</td>
<td>52 (100.0)</td>
<td>44 (88.0)</td>
</tr>
<tr>
<td>Experienced medical resuscitation</td>
<td>35 (65.4)</td>
<td>25 (50.0)</td>
</tr>
<tr>
<td>Self-assessed learner type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td>26 (50.0)</td>
<td>30 (60.0)</td>
</tr>
<tr>
<td>Auditory</td>
<td>5 (9.6)</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>Kinesthetic and tactile</td>
<td>21 (40.4)</td>
<td>17 (34.0)</td>
</tr>
</tbody>
</table>

**Table 3**

<table>
<thead>
<tr>
<th>Group Performance on Each Scale</th>
<th>Case-based Learning, Mean (SD)</th>
<th>Human Patient Simulation, Mean (SD)</th>
<th>Mean Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall score (43 items)</td>
<td>31.4 (4.1)</td>
<td>31.2 (3.6)</td>
<td>0.2 (−1.3, 1.7)</td>
</tr>
<tr>
<td>History (22 items)</td>
<td>15.9 (2.8)</td>
<td>15.5 (2.8)</td>
<td>0.4 (−0.8, 1.4)</td>
</tr>
<tr>
<td>Acute MI evaluation and management (13 items)</td>
<td>9.0 (1.9)</td>
<td>8.7 (1.9)</td>
<td>0.3 (−0.4, 1.1)</td>
</tr>
<tr>
<td>Cardiac arrest management (8 items)</td>
<td>6.5 (1.3)</td>
<td>7.0 (1.2)</td>
<td>−0.5 (−1.0, 0.02)</td>
</tr>
</tbody>
</table>

Results of the simultaneous comparison of mean subscale scores: Hotelling’s $T^2 [3,98] = 0.053, p = 0.164.$

MI = myocardial infarction.
with crossover and did not offer another training modality to the control group. While this demonstrates the ability of simulation to have a positive educational outcome, it does not address its efficacy in comparison with other teaching formats. Steadman et al. recently reported the results of a randomized controlled trial comparing acquisition of critical care skills after simulation-based training with that of problem-based learning. They concluded that simulation-based training was superior to problem-based learning in medical students; however, they acknowledged that their use of simulation as an intervention and evaluative tool could cause bias. This problem of sensitization of the simulation intervention group to the outcome measure is difficult to avoid and, as a result, is seen in a number of simulation studies. Because simulation has the ability to accurately re-create the clinical environment, it is a logical and, some would argue, superior tool for student assessment. Finding another outcome that represents clinician performance more accurately than simulation is a challenge.

To avoid sensitization of our simulation-trained group, we evaluated all participants using a chest pain and ACS OSCE. We purposely did not report students’ enjoyment or self-assessment as part of our results. The simulation literature contains ample studies demonstrating that students accept simulation as part of their medical training and believe that they learn effectively from this format. Unfortunately, student self-assessment does

---

**Figure 2.** Comparison of each item in the history subscale of the objective structured clinical examination. The dark and light lines represent the case-based learning and human patient simulation groups, respectively. Circles represent the percentage of students in each group who performed the task correctly. Hatch marks represent the 95% confidence intervals. No dark line appears for the case-based learning group for the history item of “symptom onset” because 100% of students performed this activity and thus confidence intervals could not be calculated.
OSCEs have been validated for oral certification in EM and are designed to evaluate data acquisition, problem solving, and overall clinical competence. Furthermore, in a small pilot study, performance on an oral OSCE correlated with both simulator performance and level of training. Using a unique, independent outcome measure such as the OSCE is methodologically sound and avoids bias. Our scoring sheet recorded performance of 43 different actions deemed necessary for competent evaluation and treatment of the patient with ACS. The majority of these actions require lower-order cognitive, affective, and procedural skills. The interactive learning environment of HPS is particularly effective when training cognitive strategies and situational awareness and may be less geared toward simple factual knowledge. If this is the case, the OSCE we used might miss subtle differences in higher-order areas of clinical performance. Future studies will require more focus on developing a more comprehensive outcome measure that is better tailored to the strength of the intervention. The internal consistency of the subscales presented in this study suggests that they lack evidence supporting an underlying construct that is reliably recorded, calling into question the utility of the constructs themselves. However, this does not impact the interpretation of the comparisons made between the two groups. When attempting to gain clinical expertise, the need for repetition cannot be overlooked. Ericsson links deliberate practice, defined as repetitive performance of specific skills, coupled with assessment and feedback, with improved acquisition of expertise in medicine. This definition is important. Medical education in its current form offers students limited opportunities for deliberate practice. Overtaxed faculty time cannot support the need for real-time bedside assessment and feedback. Moral obligations to prioritize patient safety restrict the opportunity for students to perform certain skills even once, let alone with the repetition necessary to achieve higher levels of performance. HPS technology provides educators with almost infinite access to a training platform for deliberate practice. Issenberg et al. demonstrated that the addition of simulation-based cardiology training to a student elective markedly improved performance in bedside cardiology skills. Our study offered students an opportunity to practice medicine in a realistic environment with expert feedback but did not allow for repetition or the opportunity to improve their performance. Participants experienced only one ACS-based scenario followed by debriefing. They did not have the

Figure 3. Comparison of each item in the acute myocardial infarction (MI) evaluation and management subscale of the objective structured clinical examination. The dark and light lines represent the case-based learning and human patient simulation groups, respectively. Circles represent the percentage of students in each group who performed the task correctly. Hatch marks represent the 95% confidence intervals.
opportunity to try different techniques and reinforce learned material. This may explain why our HPS training intervention did not result in superior performance when compared with traditional didactic training. Integrating deliberate practice into the HPS intervention in future studies is the next needed trial and may very well result in superior outcomes in the HPS intervention group. Such trials will provide the data needed to answer the question of whether all medical schools must invest in HPS to provide students with the best and most effective educational method.

LIMITATIONS

In our view, the single largest threat to the internal validity of this study design is the potential for baseline academic achievement differences in our study groups that occurred despite random assignment. We did not attempt to statistically control for previous training and patient care experiences, or for prior academic achievement, because it was believed that the risk of sensitization to the outcome of interest would in and of itself impact performance on the outcome. We therefore assume that the random assignment used created two equivalent study groups on all things measured and unmeasured. To support this assumption, we demonstrate equality between the study groups in factors that were believed to potentially affect student performance on the outcome of interest (Table 2). Nonetheless, the lack of statistical significance between the two groups on the outcome of interest may in fact be due to a starting imbalance in the academic achievement between the two groups.

Another potential threat to the internal validity of this study is history, which refers to the concurrent learning
that occurs between the intervention and the outcome evaluation. While all of the students in this study were enrolled in the same EM clerkship, the investigators were unable to control for the numbers of cardiac patients that were encountered or the numbers of codes the students witnessed before the test on the outcome measure. In addition, students studied on their own outside of prescribed reading and didactics. While unlikely, a difference in study habits between the two intervention groups could potentially mask a difference in efficacy of the educational modalities tested.

CONCLUSIONS

In this study, we compared the efficacy of HPS-based training and CBL in medical students and found no significant difference in outcomes as measured by student performance on a chest pain OSCE evaluating chest pain diagnosis and therapy. Future studies with equal methodological rigor are needed to address the issue of whether HPS produces superior educational outcomes when more extensive interventions incorporating deliberate practice are included.

References