Performance of Certified Registered Nurse Anesthetists and Anesthesiologists in a Simulation-Based Skills Assessment

Bernadette M. Henrichs, PhD, CRNA, CCRN*

Michael S. Avidan, MBBCh*

Dave J. Murray, MD*

John R. Boulet, PhD†

Joe Kras, MD*

Beverley Krause, MS, CRNA*

Rebecca Snider, RN, BSN*

Alex S. Evers, MD*

BACKGROUND: Anesthesiologists and certified registered nurse anesthetists (CRNAs) must acquire the skills to recognize and manage a variety of acute intraoperative emergencies. A simulation-based assessment provides a useful and efficient means to evaluate these skills. In this study, we evaluated and compared the performance of board-certified anesthesiologists and CRNAs managing a set of simulated intraoperative emergencies.

METHODS: We enrolled 26 CRNAs and 35 board-certified anesthesiologists in a prospective, randomized, single-blinded study. These 61 specialists each managed 8 of 12 randomly selected, scripted, intraoperative simulation exercises. Participants were expected to recognize and initiate appropriate therapy for intraoperative events during a 5-min period. Two primary raters scored 488 simulation exercises (61 participants × 8 encounters).

RESULTS: Anesthesiologists achieved a modestly higher mean overall score than CRNAs (66.6% ± 11.7 [range = 41.7%–86.7%] vs 59.9% ± 10.2 [range = 38.3%–80.4%] $P < 0.01$). There were no significant differences in performance between groups on individual encounters. The raters were consistent in their identification of key actions. The reliability of the eight-scenario assessment, with two raters for each scenario, was 0.80.

CONCLUSION: Although anesthesiologists, on average, achieved a modestly higher overall score, there was marked and similar variability in both groups. This wide range suggests that certification in either discipline may not yield uniform acumen in management of simulated intraoperative emergencies. In both groups, there were practitioners who failed to diagnose and treat simulated emergencies. If this is reflective of clinical practice, it represents a patient safety concern. Simulation-based assessment provides a tool to determine the ability of practitioners to respond appropriately to clinical emergencies. If all practitioners could effectively manage these critical events, the standard of patient care and ultimately patient safety could be improved.

(Anesth Analg 2009;108:255–62)

Anesthesia practitioners must be able to manage acute emergencies that arise during the perioperative period. In everyday practice, these emergencies are usually unforeseen and, when they do occur, patient safety must supersede any concern about evaluating a practitioner's performance and skill. Previous studies have described how simulation technology can be used to evaluate anesthesia residents' and anesthesiologists' ability to recognize the signs of a critical event and to initiate therapy to resolve crises.[1–5] In this study, we selected and designed simulated perioperative complications for events that occur infrequently (e.g., intraoperative pneumothorax) and conditions that are rarely encountered in clinical practice (e.g., malignant hyperthermia [MH]).[5] These scenarios require participants to recognize and treat an acute crisis situation, regardless of whether the situation could have been anticipated or averted by preemptive planning. With the current practice model in the United States, anesthesia care is often provided by a team comprised of anesthesiologists and certified registered nurse anesthetists (CRNAs). There are periods when CRNAs or anesthesiologists are the sole anesthesia practitioners in the operating room. There have been no published studies assessing whether there is any demonstrable differences between anesthesiologists

From the *Department of Anesthesiology, Washington University School of Medicine, St. Louis, Missouri; and †Foundation for Advancement of International Medical Education and Research (FAIMER), Philadelphia, Pennsylvania.

Accepted for publication July 28, 2008.

Supported by Division of Clinical and Translational Research, Department of Anesthesiology, Washington University (Dr. M. Avidan PI) and Foundation for Anesthesia Education and Research, Rochester, Minnesota (Dr. D. Murray PI).

Bernadette Henrichs and Michael Avidan contributed equally to this work.

Reprints will not be available from the authors.

Address correspondence to Bernadette Henrichs, PhD, CRNA, CCRN, Campus Box 8054, 660 S. Euclid Ave., St. Louis, MO 63110. Address e-mail to henrichb@anest.wustl.edu.

DOI: 10.1213/ane.0b013e31818e3d58

and CRNAs in their ability to diagnose and treat simulated emergencies in a timely manner.

Previous studies indicate that a multiple scenario format is required to reliably measure acute care skills.[4] This approach to evaluation is similar to the clinical skills assessments administered by the National Board of Medical Examiners and Educational Commission for Foreign Medical Graduates to license and certify physicians.[6,7] The goal in developing a set of intraoperative simulation exercises for anesthesia specialists is to provide reliable and valid ability estimates and to clearly demark individual as well as general areas of strength and weakness in acute care management.

Physician and nurse anesthesia educators describe similar perioperative training exercises that anesthesia residents and student CRNAs manage in simulated operating room settings.[8–15] Of particular relevance to the present investigation is the similarity between exercises designed for training CRNAs and physician anesthesiologists. This overlap suggests that a common set of skills is expected, at least in some practice domains. In a previous study, we compared the skills of anesthesia residents with those of experienced board-certified anesthesiologists using exercises created to measure acute intraoperative skills.[5] Our assessment battery included simulated intraoperative events that tap relevant and expected skills of anesthesiologists.[5] Not surprisingly, these skill sets are also expected of CRNAs. Using a similar methodology, a simulation-based evaluation can be used to measure skills expected of both CRNAs and anesthesiologists.

The study design was chosen to allow for the collection of quantitative measures of CRNA and anesthesiologist performance in managing a set of simulated intraoperative acute emergencies. Training in crisis management has historically been an informal component of both anesthesiology and nurse anesthesia training. It is possible that dissimilarities in this training have resulted in appreciably different skill levels. The purpose of this investigation was to determine whether experienced anesthesiologists and CRNAs have comparable skill levels in managing acute conditions. This detailed information about the character and range of skills that exist among physicians and nurses could be used to assess specialists' current skill levels and to develop a strategy for designing future investigations aimed at improving patient management skills.

METHODS

For this project, 12 randomly selected, scripted, intraoperative simulated events were developed to measure essential skills in intraoperative acute care management (Table 1). The 12 intraoperative events were selected from a list of intraoperative conditions created by the investigative team of anesthesia faculty and staff. The list was then rank ordered according to importance by a group of 10 anesthesiologists and 10 CRNAs from the Department of Anesthesiology at Washington University. All of the conditions selected were included in the American Board of Anesthesiology Content Outline for residents in specialty training and the Council on Accreditation of Nurse Anesthesia Educational Programs Outline for nurse anesthesia students in training.[16,17] The 12 events that were most frequently selected by both groups included (in alphabetic order) 1) acute hemorrhage, 2) anaphylaxis, 3) blocked endotracheal tube, 4) bronchospasm, 5) hyperkalemia, 6) loss of pipeline oxygen, 7) MH, 8) myocardial ischemia, 9) pneumothorax, 10) right main stem intubation, 11) total spinal, and 12) ventricular tachycardia.

Twelve scenarios were designed in a similar manner to that used for earlier studies. These scenarios were identical to those used by the same group in a previous study, in which they were shown to have good reliability and validity.[5] The scenario content required individual practitioners to recognize and manage a critical event in a 5-min time period. For many of the events, the diagnosis and treatment could be effectively accomplished in the 5-min period (mainstem intubation, pneumothorax, obstructed endotracheal tube, and bronchospasm). For conditions such as acute blood loss, anaphylaxis, hyperkalemia, MH, myocardial ischemia, total spinal, and ventricular tachycardia, participants were simply required to recognize and initiate therapy for a condition that would require continued or additional management in the period after the scenario (Table 1).

After receiving institutional review board's approval for the protocol (Washington University School of Medicine, St. Louis, MO), we obtained informed written consent from a variety of practice locations around the St. Louis area (Table 2). Six hundred solicitation letters were sent to a random selection of 300 anesthesiologists and 300 CRNAs in the St. Louis metropolitan area. It was predetermined that all respondents would be included in the study to a maximum of 50 in each group. No respondents were excluded, and 35 anesthesiologists and 26 CRNAs were enrolled. These specialists received an honorarium of $500.00 to compensate them for their time and travel during the study period. Participants were at the simulation center for approximately 4 h. Ninety-four percent of the anesthesiologists and 96% of the CRNAs worked in a team setting. The anesthesiologists recruited to participate in the study were included in a companion study designed to assess reliability and validity of the scenario content and determine whether experience in practice was related to skill level.[5]

The CRNAs and anesthesiologists who participated received a standardized orientation that included reviewing a 17-min videotape. This orientation included details about the Ohmeda® (Madison, WI) anesthesia machine, Marquette® (Piscataway, NJ) monitor, and

Table 1. Description of Events and Scoring Items for 12 Scenarios

Scenarios	Scoring items
Bronchospasm	
One minute after beginning the simulation, the oxygen saturation level decreases to 85% and heart rate increases to 120. The blood pressure remains at 105/60. Bilateral wheezing was detectable on auscultation. Elevated peak airway pressures were seen on the Bourdon gauge.	Listen to chest, increase inspired oxygen, state diagnosis, administer beta agonist/ epinephrine.
Anaphylaxis	
At 1 min, blood pressure decreases to 80/50, heart rate increases from 64 to 115, and oxygen saturation decreases to 88%. At 3 min, a skin rash clue is given.	Increase inspired oxygen, auscultate lungs, check blood pressure, state diagnosis, stop antibiotic infusion, administer epinephrine.
Unstable ventricular tachycardia	
Thirty seconds after entering simulation, wide-complex tachycardia with decreased blood pressure. Heart rate increases from 70 to 170, blood pressure decreases from 120/70 to 80/60.	State diagnosis, increase inspired oxygen, deliver shock, deliver synchronized cardioversion, give/request antiarrhythmic.
Myocardial ischemia	
Electrocardiogram indicates ST segment depression is at 3 mm. Heart rate increases to 125, and blood pressure is 170/90.	State diagnosis, check or order 12-lead electrocardiogram, administer narcotic or beta blocker therapy to decrease heart rate, request nitroglycerin drip or apply paste.
Right bronchial intubation	
At the beginning of the simulation, the vital signs are stable except the oxygen saturation is 91%. There is no chest wall movement on the left side of the chest.	Auscultation or inspection of chest, increase inspired oxygen, state diagnosis, reposition endotracheal tube.
Tension pneumothorax	
Puemothorax is present from beginning of scenario, 60 s after beginning the scenario, the blood pressure decreases to 85/55, heart rate increases to 120, oxygen saturation continues to decrease to 85%.	Auscultation of chest, increase inspired oxygen, state diagnosis, relieve with needle, or place chest tube.
Malignant hyperthermia	
Within 1 min of beginning the scenario, the heart rate and blood pressure increase to 115 and 180/90, respectively. Increased end-tidal carbon dioxide level.	State diagnosis, turn-off agent, call for Dantrolene® or malignant hyperthermia cart.
Blocked endotracheal tube	
At the beginning of the scenario vital signs are stable, except oxygen saturation is 90%. Breath sounds are distant, elevated airway pressures during controlled ventilation.	Auscultate, increase inspired oxygen, recognize increased airway pressures, pass suction catheter, state diagnosis, remove blocked endotracheal tube.
Total spinal	
The patient had a combined neuraxial/general anesthetic. Within 1 min of starting the simulation, the blood pressure starts decreasing to 60/40, heart rate decreases to 40.	Increase inspired oxygen, check blood pressure, turn-off agent, increase fluids, state diagnosis, give epinephrine.
Loss of pipeline oxygen	
Fifteen seconds after beginning the simulation, the pipeline oxygen is turned off. At 30 s, the alarm sounds. Vital signs remain stable. The first oxygen tank is empty.	State diagnosis, open oxygen tank #1, open oxygen tank #2.
Hyperkalemia	
At the beginning of the simulation, the blood pressure is 170/90, the heart rate is 75. The heart rate increases, ventricular irritability increases, and peaked *T*-waves are evident on electrocardiogram.	Order or check electrolytes or arterial blood gas or potassium, state diagnosis, institute appropriate treatment.
Acute hemorrhage	
One minute after the simulation starts the blood pressure begins to decrease to 85/50, heart rate increases to 115. One liter of "blood" is in the suction canister.	Ask about blood loss or evaluate for excessive blood loss (check suction canister), increase intravenous fluids. state diagnosis, request hemoglobin or hematocrit or blood product.

contents in the five-drawer anesthesia cart. Participants then completed an anesthesia induction using the Ohmeda (Madison, WI) machine, the MedSim Eagle® simulator, IV drugs, and airway equipment. Before beginning the individual assessment session, participants prepared their anesthesia equipment and medications in the manner they were accustomed to in clinical practice. Participants were given a copy of a preoperative assessment, an anesthesia record, and brief details about the case to review for a 10-min period before each scenario. There was no debriefing between scenarios.

Each participant managed a set of 8 of 12 possible exercises. These exercises were selected and presented in random order (determined by a random number generation). The motivation for this approach was to

Table 2. Anesthesiologists and Certified Registered Nurse Anesthetists—Demographic Profile

	Practicing physicians	CRNAs
Mean age	43 (33–57)	45.4 (38–55)
Number of females/male	6/29	15/11
Practice setting: community	22	20
Practice setting: teaching	9	4
Average years in practice	11.9 (2–26)	11.1 (3–32)
Type of practice: team	26	24
Type of practice: individual	6	1
Type of practice: both team and individual	3	1

prevent participants from disseminating information about the scenarios. Participants were told that the same scenarios would not be duplicated from participant to participant. The same random number assignment was used for each practitioner group in an effort to assure similar scenario exposure in the anesthesiologist and CRNAs groups. Based on this protocol, the groups were exposed equally to each of the exercises ($\chi^2_{11} = 1.25$, $P = 1.00$). Therefore, it is reasonable to conclude that the two practitioner groups received, on average, assessments of equivalent scope and difficulty.

The simulated event occurred approximately 20–30 min after induction. The participants were instructed to assume care for the mannequin after entry into the simulation laboratory. No one else was present in the room. A faculty member was in an adjacent control room behind a one-way mirror. Participants were given clear instructions to state what the diagnosis of the emergency was and what treatments they were instituting. The faculty members conducted the simulation and ensured that the vital signs of the mannequin responded appropriately according to the interventions of the participants. The initial simulated signs matched the anesthetic record. The intraoperative crisis and summary of the signs present and scoring measures are presented in Table 1. Audiovisual recordings were made of each scenario to facilitate scoring and to allow independent review and further analyses.

Two trained raters, a research nurse and a physician, who were blinded to the identity and prior training of the participants, reviewed the videotaped performances. They were formally trained in assessing each of the scenarios and were given specific instructions on scoring. The raters were not anesthesia practitioners and did not know any of the participants. If the two raters disagreed or had differences of more than 30 s regarding when the action occurred, a third rater reviewed the scenario and provided scores. For scenarios with scoring discrepancies, the participant was given the score for the key action based on modal rating (where two of the three raters agreed that the action was or was not completed). Although timing data were collected, they were not explicitly used in the scoring process. Instead, a point was given if the key action was completed any time in the 5-min period; otherwise, no point was given. The score on each exercise was based on the sum of the key actions. The number of key actions ranged from 3 to 6, depending on the specific scenario. In addition, a percent score based on the number of key actions credited out of the possible key actions was computed. A total score for each participant was computed as the average of percentage encounter scores across the eight scenarios that he/she encountered.

To investigate potential performance differences, by group, a two-way analysis of variance (ANOVA) was used. The dependent variable was the key action score. The independent variables were encounter and group (practicing physician, CRNA). For any significant interactions and main effects, appropriate follow-up analyses (e.g., one-way ANOVA) were used. To provide a measure of scenario discrimination, the overall assessment scores (the average of each participant's eight scenario scores) were correlated with each of the 12 individual scenario scores. This provides a measure of whether performance on an individual scenario can be used to predict overall performance.

The overall reliability of the participant scores (both anesthesiologists and CRNAs combined) were determined using the generalizability theory.[18,19] This analysis is used to estimate and quantify the sources of variability in the participant scores, including those connected with the raters, the specific scenarios, the participants, and the associated interactions. Interrater reliability was estimated by correlating the key action scores for two primary raters. This was performed overall and by encounter. Descriptive statistics (means and standard deviation) were used to summarize participant performance, by scenario.

RESULTS

Thirty-five anesthesiologists and 26 CRNAs managed 8 of 12 exercises during an individual assessment session. Two primary raters scored 488 simulation exercises (61 participants × 8 simulated events) (Table 3). The results for the performances of the anesthesiologists are also reported in a companion manuscript.[5]

The two-way ANOVA yielded a significant group effect ($F_1 = 7.8$, $P < 0.01$). The significant group effect indicates that, averaged over their eight encounters, the anesthesiologists (mean 66.6% ± 11.7; range = 41.7%–86.7%) received higher overall scores than the CRNAs (59.9% ± 10.2; range = 38.3%–80.4%) (Fig. 1). There was no significant group by scenario interaction suggesting that the overall difference in performance, by group, was reasonably consistent from scenario to scenario (Fig. 2). There was a significant effect attributable to scenario ($F_{11} = 60.7$, $P < 0.01$). This indicates that the combined average of anesthesiologists' and CRNAs' mean scores varied considerably as a function of the clinical content of the scenario (Fig. 2). Performance did not improve as individuals progressed through each of the eight scenarios.

Table 3. Key-Action Scores/Scenario Discrimination for Certified Registered Nurse Anesthetists (CRNAs) and Anesthesiologists

Individual scenarios (n = 488)	No. of key actions in scenario	CRNA key action score, mean ± SD (n)	Anesthesiologist key action score mean ± SD (n)	Scenario discrimination (case-total correlation)
Bronchospasm (participants = 38)	4	3.73 ± 0.46 (n = 15)	3.78 ± 0.42 (n = 23)	0.26
Anaphylaxis (participants = 38)	6	3.76 ± 0.97 (n = 17)	3.81 ± 1.21 (n = 21)	0.34
Vent tachycardia (participants = 41)	5	3.89 ± 0.81 (n = 19)	3.73 ± 1.03 (n = 22)	0.52
Myocardial ischemia (participants = 43)	4	3.18 ± 1.59 (n = 17)	3.69 ± 1.33 (n = 26)	0.65
Right main intubation (participants = 41)	4	3.71 ± 0.47 (n = 17)	3.63 ± 0.58 (n = 24)	0.20
Tension pneumothorax (participants = 39)	4	3.94 ± 1.00 (n = 16)	4.17 ± 1.03 (n = 23)	0.28
Malignant hyperthermia (participants = 41)	3	0.31 ± 0.87 (n = 19)	0.56 ± 1.00 (n = 22)	0.27
Blocked endotracheal tube (participants = 43)	6	2.35 ± 1.06 (n = 17)	3.04 ± 1.18 (n = 26)	0.49
Total spinal (participants = 41)	6	2.82 ± 1.01 (n = 17)	3.30 ± 1.18 (n = 24)	0.43
O_2 failure (participants = 42)	3	2.66 ± 0.47 (n = 19)	2.70 ± 0.63 (n = 23)	0.40
Hyperkalemia (participants = 41)	3	0.11 ± 0.32 (n = 18)	0.52 ± 0.85 (n = 23)	0.48
Acute hemorrhage (participants = 40)	4	1.94 ± 1.48 (n = 17)	2.57 ± 1.44 (n = 23)	0.65
Total no. of scenarios		2.70 ± 0.65 (n = 208)	3.02 ± 0.57 (n = 280)	

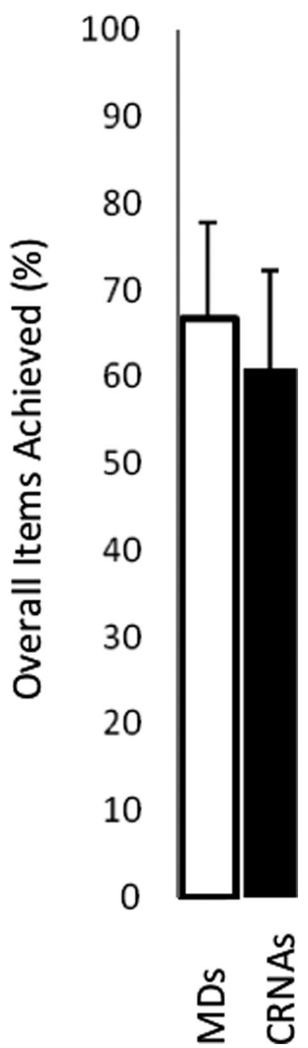

Figure 1. Overall scores. The columns represent the mean for the overall scores of all eight scenarios by individuals in both groups. The error bars represent standard deviations.

The generalizability coefficient (reliability estimate) for the simulation scores, based on eight scenarios and two raters, was 0.80. The main sources of variance in participant scores could be attributed to differences in scenario difficulty and to the resultant fluctuations in individual performance from one encounter to the next (task sampling variability). The variation in participants mean score among the scenarios is illustrated in the percentage of actions completed in Figure 2. The majority of participants achieved more than 80% of the key action items for bronchospasm, right mainstem intubation, pneumothorax, and loss of pipeline oxygen (Fig. 2). In contrast, fewer than half the participants were able to accomplish any of the key actions in the MH (mean key action score = 0.44) or the hyperkalemia (mean key action score = 0.36) scenarios. These scenarios also produced some of the most variable key action scores, with 0% (no key items credited) and 100% (all items credited) being observed among participants. The scenario difficulties (number of key items attained) and scenario discriminations (scenario-total correlations) are presented in Table 3. The discrimination values ranged from a low of 0.20 (right mainstem intubation) to 0.65 (acute hemorrhage and myocardial ischemia). All values were positive indicating that the modeled scenarios are effective in differentiating between participants of low and high ability.

As part of the variance components analysis described earlier, we found that the choice of rater, and associated rater interactions, had little impact on the reproducibility of the simulation scores. As a result, interrater reliability, based on the 488 double scored encounters, was quite high (r = 0.88). Interrater reliability was lowest for loss of pipeline oxygen (r = 0.38) and highest for MH (r = 0.98). Using only the two primary raters, and comparing their individual key

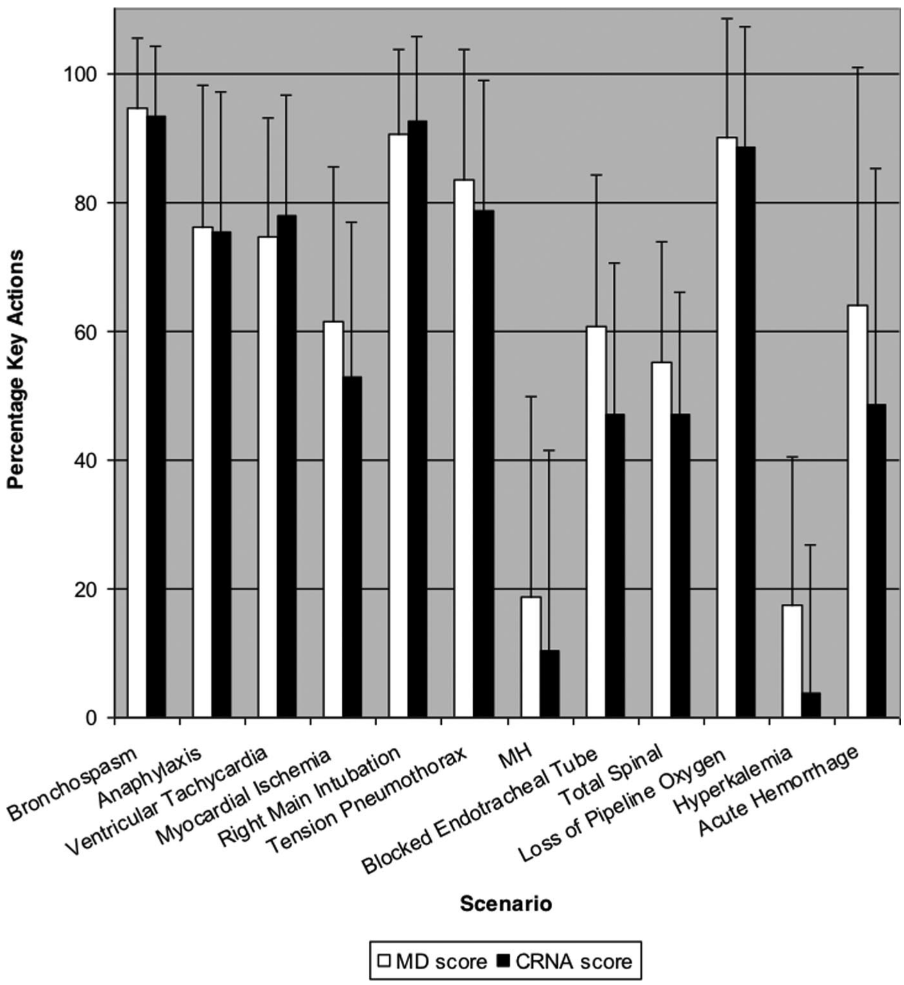

Figure 2. Percentage of key actions accomplished. The columns indicate the mean percentage of key actions accomplished by anesthesiologists (MD) and certified registered nurse Anesthetists (CRNA) groups for each scenario. The error bars represent standard deviations.

action scoring, the scenario discrepancies ranged from one disagreement (bronchospasm, $n = 38$ participants, 152 scored actions, $n = 4$ key actions) to 10 disagreements (acute hemorrhage, $n = 40$ participants, 160 scored actions, $n = 4$ key actions). Overall, there were only 66 total instances in which documented rater discrepancies for key actions required the review by a third expert.

DISCUSSION

On average, the anesthesiologists received modestly higher scores than the CRNAs. This overall difference, although statistically significant, was small in magnitude. There were individuals in both groups who, across most of the scenarios, made the correct diagnoses and instituted appropriate treatment within the 5-min time limit. Anesthesiologists and CRNAs were variably adept in the initial diagnosis and management of simulated intraoperative emergencies. When an emergency situation arises, our study suggests that some anesthesiologists and CRNAs could effectively recognize the condition and initiate therapy in a simulation. Worryingly, our study also suggests that some may not respond effectively.

This study exposes a potential intraoperative patient care concern. In both groups, there were individual anesthesia practitioners who in many instances failed to diagnose and treat appropriately simulated, life-threatening emergencies within a 5-min period. This suggests that if the 5-min scenarios used in this study represent performance in real clinical care, there may be a need for continuing education and improvement of anesthesia practitioners' abilities to diagnose and manage simulated life-threatening emergencies. Patient simulators may be useful in assessing practitioners' skills in this regard and, perhaps more importantly, for providing training in the management of both rare and common intraoperative emergencies that may directly result in improving patient safety.

CRNAs and anesthesiologists achieved maximum scores on exercises, such as the bronchospasm and loss of pipeline oxygen, thus providing evidence that the required skill sets are common and well managed by both practitioner groups. Participants had difficulty in recognizing and managing MH and hyperkalemia exercises. The scores varied on these difficult exercises, but those who effectively recognized and managed these events were likely to obtain higher overall scores. This suggests that the content of the MH and hyperkalemia events, although clearly more difficult and challenging, provides useful information to discriminate between low- and high-ability practitioners. Follow-up investigations that determine specialty-specific performance expectations for these events would indicate whether the participants' scores represent a skill deficit in recognizing these conditions or, alternatively, that the design, content, or scoring of these scenarios was not appropriate.

The fact that there was a wide range of performance in both groups suggests that certification and experience as either a CRNA or as an anesthesiologist does not lead to a common, defined practice standard for this type of skill measure within a specialty. Most experts, including those recruited from our programs, agree that the selected exercises represent conditions that are essential to anesthesia practice.[20] Our findings support previous studies that have reported performance shortfalls among board-certified anesthesiologists in academic and community practice.[9–13] It is notable, however, that the scores measured in this study were worse than the technical scores measured in another study.[8]

If targeted remediation of practitioners who received lower scores, especially for specific types of scenarios, was available, then practice standards could be elevated.

Going forward, follow-up studies could be used to establish skill-based competency standards for anesthesia practice. Ultimately, this approach could lead to one of the first specialty-specific skills courses, akin to the widely applied advanced cardiac life support or pediatric advanced life support skills training programs, in which abilities can be measured and compared with defined performance standards.

Although we were able to reliably evaluate acute care skills with a multistation assessment, there are a number of limitations of our findings. First, relatively few, self-selected practitioners agreed to participate. To the extent that these individuals are not representative of the population of anesthesiologists and CRNAs, our results may not be generalizable. Second, although the participants were compensated for their time, factors such as motivation, assessment anxiety, familiarity with the simulator, and anesthesia equipment may have affected their performance. As a result, there may be some systematic error in our derived ability estimates. Third, our study focused more on technical skills, individual performance, and crisis management than on core knowledge and teamwork. Fourth, the practitioners who participated in this study overwhelmingly work in team-based practices; skill sets of solo practitioners may be different. Fifth, the 5-min scenario represents an artificial construct, which is different from the actual way anesthesia is delivered. Finally, the findings are based on performance in a simulated environment. Even though the fidelity is quite good, performance with real patients, especially when there are other health care practitioners present, may be better or worse. To address these limitations, studies involving larger groups of specialists, including those from more diverse geographic areas and practice settings, are required.

This study was intentionally limited in its scope. The methodology used simulated events that were selected and designed to measure and compare skills that are common to both groups rather than address the entire range of skills expected of CRNAs and anesthesiologists. Anesthesia practitioners have a variety of skills and roles, ranging from preoperative assessment to postoperative care. In most situations, anesthesiologists and CRNAs have clearly demarcated roles and functions and work together as a team. Therefore, assessing individual practitioners will not fully capture some aspects of patient care, including team work and professionalism. Moreover, the results may not generalize to areas outside of the intraoperative emergency domain.

This study provides a methodology for achieving a reproducible measure of acute care skills. The reliability estimate of 0.80 is similar to that achieved for performance-based examinations of comparable length.[21,22] This level of consistency was achieved by using eight acute exercises, a key action scoring method, and two independent raters. From a measurement perspective, we found that choice of rater contributed relatively little to the precision of the scores. This finding is consistent with what has been reported in the performance assessment domain.[19,21]

In this simulation-based skill assessment, anesthesiologists achieved modestly higher overall scores than CRNAs. However, the similar broad range in both groups indicates that certification and clinical practice alone are not sufficient guarantees of universal satisfactory performance. Systematic crisis skill training with a high-fidelity simulator or with other methods should be incorporated into the formal curriculum of both anesthesiology and CRNA trainees. Clinical simulation can provide anesthesia practitioners with useful training in the recognition and management of life-threatening intraoperative emergencies. If all practitioners could effectively manage these critical events, the standard of patient care and ultimately patient safety could be improved.

REFERENCES

1. Murray DJ, Boulet J, Ziv A, Kras J, McAllister JD, Woodhouse J. An acute care skills evaluation for graduating medical students: a pilot study using clinical simulation. Med Educ 2002;36: 833–41
2. Murray DJ, Boulet J, Kras JF, Woodhouse JA, Cox T, McAllister JD. Acute care skills in anesthesia practice: a simulation-based resident performance assessment. Anesthesiology 2004;101: 1085–94
3. Murray DJ, Boulet JR, Kras JF, Cox T, McAllister JD. A simulation-based acute skills performance assessment for anesthesia training. Anesth Analg 2005;101:1127–34
4. Boulet JR, Murray DJ, Kras JF, Woodhouse J, McAllister JD, Ziv A. Assessing the acute-care skills of medical students and recent graduates: reliability and validity of patient simulator scores. Anesthesiology 2003;99:1270–80
5. Murray DJ, Boulet J, Avidan MS, Kras JF, Henrichs BM, Woodhouse JA, Evers AS. Performance of residents and anesthesiologists in a simulation-based skills assessment. Anesthesiology 2007;107:705–13

6. United States Medical Licensing Examination™ (USMLE™) Step 3. Available at: www.usmle.org. Federation of State Medical Boards (FSMB) and National Board of Medical Examiners® (NBME®) Philadelphia, PA
7. Dillon GF, Boulet JR, Hawkins RE, Swanson DB. Simulations in the United States Medical Licensing Examination™ (USMLE™). Qual Saf Health Care 2004;13(suppl 1):i41–i45
8. Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. Anesthesiology 1998;89:8–18
9. DeAnda A, Gaba D. The role of experience in the response to simulated critical incidents. Anesth Analg 1991;72:308–15
10. Jacobsen J, Lindekaer AL, Ostergaard HT, Nielsen K, Ostergaard D, Laub M, Jensen PF, Johannessen N. Management of anaphylactic shock using a full scale anaesthesia simulator. Acta Anaesth Scan 2001;45:315–9
11. Lindekaer AL, Jacobsen J, Andersen G, Laub M, Jensen PF. Treatment of ventricular fibrillation during anaesthesia in an anaesthesia simulator. Acta Anaesthesiol Scand 1997;41:1280–4
12. Monti EJ, Wren K, Haas R, Lupien AE. The use of an anesthesia simulator in graduate and undergraduate education. CRNA 1998;9:59–66
13. Schwid HA, O'Donnell D. Anesthesiologists' management of simulated critical incidents. Anesthesiology 1992;76:495–501
14. Schwid HA, Rooke GA, Carline J, Steadman RH, Murray WB, Olympio M, Tarver S, Steckner K, Wetstone S. Evaluation of anesthesia residents using mannequin-based simulation: a multi-institutional study. Anesthesiology 2002;97:1434–44
15. O'Donnell J, Fletcher J, Dixon B, Palmer L. Planning and implementing an anesthesia crisis resource management course for student nurse anesthetists. CRNA 1998;9:50–8
16. Content Outline for Residents in Specialty Training. Joint Council on In-training Examinations. Chicago: ASA information, 2003
17. Council on Accreditation of Nurse Anesthesia Educational Programs. Content outline. Chicago: COA brochure, 2003
18. Brennan RL, Gao X. Variability of estimated variance components and related statistics in a performance assessment. Appl Meas Educ 2001;14:9–12
19. Boulet J. Generalizability theory: basics. In: Everitt BS, Howell DC, eds. Encyclopedia of statistics in behavioral science. Chichester, England: Wiley, 2005:704–11
20. Gaba DM. What makes a good anesthesiologist? Anesthesiology 2004;101:1061–3
21. Boulet J, Rebbecchi T, Denton EC, McKinley DW, Whelan GP. Assessing the written communication skills of medical school graduates. Adv Health Sci Educ 2004;9:47–60
22. Gimpel JR, Boulet JR, Errichetti AM. Evaluating the clinical skills of osteopathic medical students. J Am Osteopath Assoc 2003;103:267–79