INFLUENCE OF PULSE OXIMETRY AND CAPNOGRAPHY ON TIME TO DIAGNOSIS OF CRITICAL INCIDENTS IN ANESTHESIA: A PILOT STUDY USING A FULL-SCALE PATIENT SIMULATOR*

Samsun Lampotang, PhD,[1–4]
J. S. Gravenstein, MD,[1] *Tammy Y. Euliano, MD,*[1]
Willem L. van Meurs, PhD,[1, 2, 4]
Michael L. Good, MD,[1, 2, 5] *Paul Kubilis, MS,*[6]
and Rod Westhorpe, FRCA, FANZCA[7]

Lampotang S, Gravenstein JS, Euliano TY, van Meurs WL, Good ML, Kubilis P, Westhorpe R. Influence of pulse oximetry and capnography on time to diagnosis of critical incidents in anesthesia: A pilot study using a full-scale patient simulator.
J Clin Monit 1998; 14: 313–321

ABSTRACT. Objective. Many studies (outcome, epidemiological) have tested the hypothesis that pulse oximetry and capnography affect the outcome of anesthetic care. Uncontrollable variables in clinical studies make it difficult to generate statistically conclusive data. In the present study, we eliminated the variability among patients and operative procedures by using a full-scale patient simulator. We tested the hypothesis that pulse oximetry and capnography shorten the time to diagnosis of critical incidents. **Methods.** A simulator was programmed to represent a patient undergoing medullary nailing of a fractured femur under general anesthesia and suffering either malignant hyperthermia, a pneumothorax, a pulmonary embolism or an anoxic oxygen supply. One hundred thirteen anesthesiologists were randomly assigned to one of two groups of equal size, one with access to pulse oximetry and capnography data and the other without. Each anesthesiologist was further randomized to one of the four critical incidents. Each anesthetic procedure was videotaped. The time to correct diagnosis was measured and analyzed. **Results.** Based on analysis of 91 of the subjects, time to diagnosis was significantly shorter (median of 432 s vs. >480 s) for the anoxic oxygen supply scenario ($p = 0.019$) with pulse oximetry and capnography than without. No statistical difference in time to diagnosis was obtained between groups for the other three critical incidents. **Conclusions.** Simulation may offer new approaches to the study of monitoring technology. However, the limitations of current simulators and the resources required to perform simulator-based research are impediments to wide-spread use of this tool.

KEY WORDS. Pulse oximetry, capnography, monitoring, critical incidents, simulator, anoxic oxygen supply, pneumothorax, pulmonary embolism, malignant hyperthermia.

From the [1]Department of Anesthesiology, University of Florida College of Medicine, [2]University of Florida Brain Institute, [3]Department of Mechanical Engineering, University of Florida College of Engineering, [4]Department of Electrical and Computer Engineering, University of Florida College of Engineering, [5]Veterans Affairs Medical Center, Gainesville, Florida; [6]Division of BioStatistics, University of Florida, [7]The Royal Children's Hospital, Parkville, Melbourne, Australia.

Received Oct 6, 1997, and in revised form May 12, 1998. Accepted for publication Apr 28, 1998.

Address correspondence to Samsun Lampotang, PhD, c/o The Editorial Office, University of Florida College of Medicine, Department of Anesthesiology, P.O. Box 100254, Gainesville, FL 32610-0254, E-mail: sem@anest4.anest.ufl.edu

* The University of Florida owns and licenses human patient simulator technology. Royalties received by the University of Florida are distributed in part to the inventing team, which includes Drs Euliano, Good, Gravenstein, Lampotang, and van Meurs.

INTRODUCTION

Several investigators have tested the hypothesis that pulse oximetry [1–3] and capnography [4] extend measurable and clinically significant benefits. While these studies confirm that the technology does indeed provide interesting clinical information, they have failed to generate statistically valid data attesting to a reduction in preventable anesthetic mortality. Full-scale patient simulators offer reproducible clinical scenarios, thus eliminating the many confounding factors inherent to clinical studies such as uncontrollable differences among patients and procedures. Using a simulator representing a single patient experiencing one of four critical incidents during general anesthesia, we tested the hypothesis that

pulse oximetry and capnography can shorten the time to arrive at the correct diagnosis.

METHODS

We conducted the study in April 1996, in the exhibition hall of the 11th World Congress of Anaesthesiologists in Sydney, Australia (WCA 96). The logistics of conducting a simulator study during a scientific meeting are described in a separate paper [5].

Two anesthesiologists from the University of Florida (UF) and 13 anaesthetists from Australia and New Zealand were the simulator instructors (see Acknowledgments section), supported by 6 UF simulator team members. The UF team developed the scenarios used for the study. On the day preceding the study, all instructors received two hours of on site training from the UF simulator instructors.

For our study, we used a full-scale patient simulator consisting of a model-driven, script-controlled, life-sized mannequin (Human Patient Simulator – HPS, Medical Education Technologies, Inc., Sarasota, FL) [6–8].

Each scenario script is implemented as a finite state algorithm (a.k.a. a finite state machine) which is basically a sequence of defined clinical states and the factors or events that determine transition from one state to another. Within each finite state, simulation parameters are specified, e.g., body temperature and neuromuscular blockade. As each scenario unfolds, transitions from one finite state to another occur and physiologic and pharmacologic models alter the course of the simulation based on the scripted changes and/or the interventions of the user. For the pneumothorax script, we did not program changes in end-tidal PCO_2 or SpO_2 but merely prevented one lung from expanding (the left intrapleural volume was increased by 3 l to simulate left lung collapse, resulting in raised peak inspiratory pressure) and increased the shunt fraction of the respiratory model. As a consequence, changes in breath sounds and peak inspiratory pressure preceded changes in end-tidal PCO_2 and SpO_2. For the MH incident, we increased oxygen consumption which raised CO_2 production (reflected in end-tidal PCO_2). Simultaneously we increased arterial pressure and heart rate. For the anoxic oxygen supply scenario, we substituted nitrogen for oxygen in the gas supply and then let the physiologic models govern all physiologic responses; we had to decrease the sensitivity of the heart to ischemia in order to delay the onset of ventricular tachycardia and fibrillation. For the pulmonary embolism event, we increased pulmonary vascular resistance and the alveolar dead space and allowed the physiologic models to respond to these changes.

All participants in the study completed the following steps: (1) a 15 minute introductory session to the anesthesia machine and monitoring equipment, (2) a 15 minute introductory session to the patient simulator, (3) a 15 minute videotaped session using the simulator during one of the scenarios and (4) a 15 minute debriefing session using videotape playback and printed time plots of physiological variables exported by the HPS (Figure 1). The focus of the study was step 3 during which the participant confronted the critical incident.

We programmed the patient simulator to represent the case history (Figure 2) which was given to the participant before he or she was asked to assume responsibility for the patient.

The participant was then invited to assume the care of this patient. The patient's lungs were mechanically ventilated (exhaled tidal volume 750 ml, RR 10 breaths/min, inspiratory : expiratory time ratio 1 : 2, fresh gas flow consisted of 1.0 l/min O_2 and 1.6 l/min "N_2O," actually N_2 because scavenging was not available). The capnograph and pulse oximeter probes were already applied to the patient simulator.

However, for 50% of the participants, the area on the physiological monitor (SpaceLabs, Redmond, WA) where the pulse oximetry and capnography data were displayed was masked with the pulse oximeter tones turned off while the same data, including pulse oximeter tones, were available for the remaining 50% of the participants. We turned off the alarms for FiO_2, SpO_2, CO_2, NIBP and ECG for the duration of the study. A self-inflating resuscitation bag was available.

During the first two minutes of the simulator session, the participants were invited to acquaint themselves with the patient simulator and the clinical setting, establish baselines of physiologic variables, auscultate the lungs, review the available drugs, check the anesthesia machine and ask questions. Then the participant took over the anesthetic.

Once the participant had assumed care of the patient, we triggered one of the four critical incidents, based on a randomized scenario sequence prepared ahead of time. The participant was unaware that only four critical incidents were being simulated. The changes in the measured variables for the anoxic oxygen supply incident, without clinical intervention, are shown in Figure 1. Regardless of the scenario, the case history and the simulation contained specific events providing identical historical clues for the 4 possible diagnoses: for malignant hyperthermia, the patient had been given succinylcholine; for pulmonary embolism, the surgeon was in the process of inserting an intramedullary rod

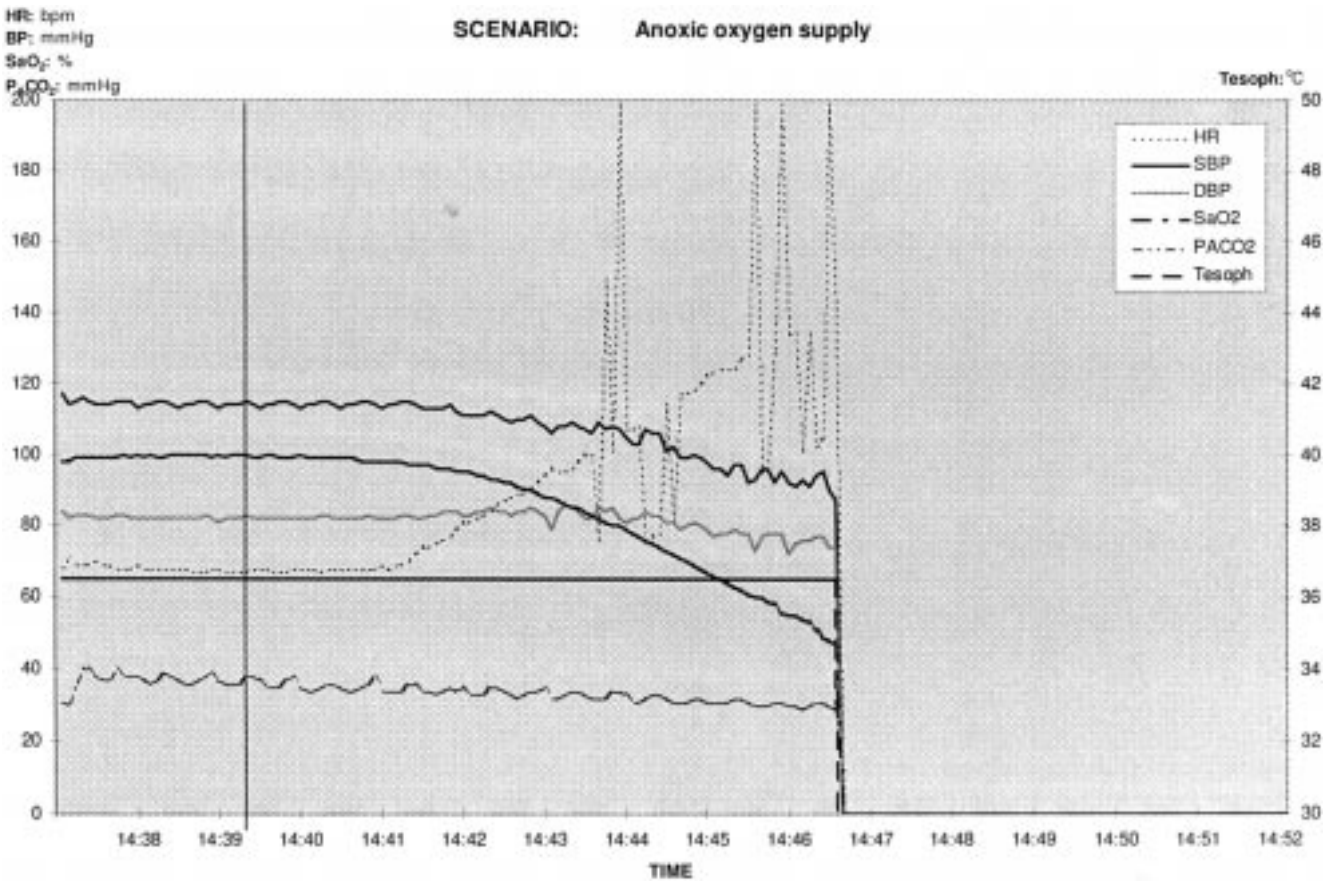

Fig. 1. The time plot of heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), oxygen saturation (SaO_2), alveolar partial pressure of carbon dioxide ($PaCO_2$) and esophageal temperature (Tesoph) during an anoxic oxygen supply scenario, without clinical intervention. The spikes in the heart rate are the result of arrhythmias. Time is indicated on the horizontal axis as actual time of day in hour and minutes. Plots similar to Figure 1 for the other 3 critical incidents can be viewed at http://gasnet.med.yale.edu/periodical/jcmc/1998/July.

into the femur; for pneumothorax, the patient had a history of chest trauma; for anoxic oxygen supply, at the beginning of *every* scenario a helper appeared with an oxygen E-size cylinder, busied himself behind the anesthesia machine and clanked the cylinders to indicate an exchange, mentioned his intervention and then showed the exchanged oxygen cylinder to the anaesthetist.

In each simulator session, only one critical event was simulated. Table 1 shows the clinical cues that enabled a diagnosis for each critical incident. We group-randomized the sequence of scenarios in order to minimize potential operator bias and to ensure equal representation of the four critical incidents, with and without pulse oximetry and capnography data. For the purpose of group randomization, each critical incident effectively becomes two scenarios, depending on whether the critical incident is simulated with, or without, the availability of SpO_2 and CO_2 data, thus producing a total of 8 scenarios.

Because we did not simulate the surgery or the surgeon, the instructors gave clinical information that would normally have been obtained from the surgical field; for example, mentioning that an intramedullary rod was being inserted into the femur, that the blood was dusky or blue whenever that was expected, that bleeding was not excessive. They also answered questions when asked, for example that the oxygen analyzer in the breathing circuit was broken, that arterial blood gases were available, that a urinary catheter was not in place and that blood or intravenous fluids were available. If the participant took actions without enunciating a diagnosis, the instructor inquired for the reason of the action.

Each simulator exercise was scheduled to run for up to 15 minutes. If the participant had not arrived at a diagnosis within 8 minutes after the critical incident

"The patient is a 65 year old, 70 kg farmer who was hit by a bus. He has sustained bruises and abrasions over his chest and abdomen and a compound fracture of the left femur.

The patient claims not to have been sick a day in his life but admits to shortness of breath with strenuous work.

In the emergency room he has complained of pain in his leg and tenderness over the chest and abdomen. On physical examination he appears healthy, his vital signs are stable, he is tender to palpation over the chest and abdomen. Breath sounds are equal bilaterally. He is breathing without difficulty but complains of left sided chest pain when taking a deep breath. BP 120/85 mm Hg, HR 79 beats/min, respiratory rate is 24 breaths/min, temperature is 36.9 °C, ECG is normal. No other history is available.

A rapid sequence induction using thiopental 400 mg and succinylcholine 100 mg followed by intubation, with cricoid pressure, was easy. He has been given 60 mg morphine sulfate intravenously in divided doses, pancuronium 8 mg, and now receives 70% nitrous oxide in oxygen. His lungs are mechanically ventilated.

The operation has been under way for ½ hour. The anesthetist has developed abdominal cramps and diarrhea and must leave the operating theater. You are to take over the anesthetic.

Monitors in place include a non-invasive blood pressure (NIBP) cuff that cycles every 2 minutes, an ECG with leads II and V5 and an esophageal temperature probe. Some anaesthetists will also have a capnograph and pulse oximeter in place. Gas for the anesthesia machine is supplied by E cylinders.

The surgeon is in the process of internal fixation of the fractured femur.

The anesthetic record shows the course of the anesthetic to have been uncomplicated. There are no urgent issues to be addressed. Fluid replacement has so far been 3100 ml Ringer's lactate solution. No urinary catheter is available."

Fig. 2. The case stem used in the study.

was triggered, the instructor helped the participant to make a diagnosis so that no patient died.

We video- and audiotaped the simulator sessions using a wide-angle video camera that enabled review of all the steps taken and comments made by the participating clinician and the attending instructor who both wore microphones. We connected two display screens to the physiological monitor. We placed one display screen at its usual location on the anesthesia machine and masked its SpO_2 and CO_2 data when the randomized scenario sequence called for it. The second display screen, showing identical data to the first display, was never masked and oriented away from the participant's view. A second videocamera took a close-up shot of the second display which we superimposed over a corner of the wide-angle video shot ("picture in a picture"). Thus, the physiologic data (ECG, non-invasive blood pressure, temperature, pulse oximeter, and capnograph) captured by the physiological monitor were always available for the ensuing debriefing and the subsequent data analysis, whether or not the clinician had data from the pulse oximeter and capnograph available during his or her simulation session.

All relevant physiologic data (e.g., heart rate, systolic and diastolic blood pressures, $PaCO_2$, SpO_2, esophageal temperature, end-tidal PCO_2) and internal parameters of the patient simulator as well as all drugs and fluids administered during the exercise by the clinician were recorded every 5 seconds to a data file. We copied the data file to a computer diskette at the conclusion of the simulator session for use during the ensuing debriefing session.

The participants were asked to fill out a questionnaire about their training in anesthesiology and current practice. The questionnaire was subsequently analyzed to verify appropriate randomization of the participants and that the populations in the two groups (with and without pulse oximetry and capnography) were comparable overall, as well as for each of the 4 scenarios.

All videotapes were reviewed and transcribed off-line by a single experienced anesthesiologist (JSG) who observed the participants' actions and comments, thus avoiding inter-observer variability. The observer was not blinded to the nature of the critical incident or availability of SpO_2 and CO_2 data. The time to diagnosis interval started with the first observable indication of a critical incident on the monitor that was out of the participant's view. We started the stopwatch with the first sustained change in heart rate for pneumothorax and pulmonary embolism, a sustained rise in end-tidal PCO_2 for malignant hyperthermia and a sustained drop in SpO_2 for the anoxic oxygen supply scenario. The time to diagnosis ended with the participant stating the diagnosis or performing the appropriate corrective action. If no diagnosis had been reached after 8 minutes, that fact was recorded.

We targeted a minimum sample size of 100 subjects for the pilot study, based on the logistics of the simulation exercise and the meeting days available.

Statistical methods

The statistical analysis consisted of three parts: (a) confirmation that the demographic data for monitoring and control groups were not significantly different, (b) analysis of the pilot data to determine if statistically significant differences in diagnosis time were obtained and (c) a retrospective power analysis to determine the theoretical sample size required to establish a clinically relevant significant difference between the groups with and without SpO_2 and CO_2 monitoring.

Table 1. The pattern of vital sign changes associated with each critical incident

Scenario	Breath sounds	Peak inspiratory pressure	Heart rate	Blood pressure	Temperature	SpO_2	End-tidal CO_2
Anoxic oxygen supply	L = R	Normal	↑	↓	Normal	↓	Normal
Pneumothorax	L ≠ R	↑	↑	↓	Normal	↓	↓
Pulmonary embolism	L = R	Normal	↑	↓	Normal	↓	↓
Malignant hyperthermia	L = R	Normal	↑	↑↓	↑	Normal then ↓	↑

In order to assess subgroup balance within each critical incident group, demographic characteristic means, rank sums, or proportions were compared between monitoring and non-monitoring subgroups using respectively the independent-sample t-test (age, years of training, years of practice), the Wilcoxon rank sum test (level of previous experience with pulse oximeter or capnograph monitoring), or the Fisher exact probability test (sex, practice location) [9]. The software package used for the statistical analysis was SAS, version 6.11 (SAS Institute, Cary, NC).

The time taken to correctly diagnose the critical incident was recorded for each anesthesiologist. Correct diagnosis times greater than 8 minutes after triggering the critical incident were considered right-censored at 8 minutes. For each critical incident group, Kaplan–Meier survival curves indicating the probability of making a correct diagnosis as a function of time were constructed for the monitoring and non-monitoring subgroups for each critical incident and compared using the log-rank test. Cox proportional hazards regression was used to estimate hazard ratios (or relative risks) with 95% confidence bounds for the probability of making a correct diagnosis in the monitoring subgroup relative to the non-monitoring subgroup. Estimates of the median time to a correct diagnosis with 95% confidence bounds were also calculated for each subgroup when possible.

The time to diagnosis data were also analyzed in a "binary" fashion, i.e., did the participant correctly identify the incident within 8 minutes or not? The proportion of anaesthetists in each subgroup making a correct diagnosis within 8 minutes was compared between subgroups using the Fisher exact probability test (a.k.a., Fisher exact test). The Cox regression and the exact odds ratio were evaluated using the Egret Epidemiological Statistics software package (Statistics and Epidemiology Research Corporation, Seattle, WA).

Assuming exponentially distributed diagnosis times and equal monitor and control group sample sizes, total study sample size required to detect selected minimum percent reductions in the mean diagnosis time relative to control at 80% power and a significance level of 0.05 was computed for various combinations of control mean diagnosis time and simulation interval length. For the power analysis, after consulting with clinical anesthesiologists, we assumed that a 10% reduction in diagnosis time was clinically relevant. The power analysis was performed using the Egret SIZ module (Statistics and Epidemiology Research Corporation, Seattle, WA).

RESULTS

During 5 consecutive days, 113 anesthesiologists attending the WCA 96 meeting participated in the study. Four of these were excluded on site because of protocol violations, leaving 109 valid subjects for data analysis. During shipment of the videotapes from Australia to the United States, one box of 18 videotapes representing 18 subjects was lost leaving us with 91 participants to analyze.

With one exception, no significant demographic differences were observed between the monitoring and non-monitoring subgroups in any of the critical incident groups. The exception was in the MH critical incident group, where the proportion of anesthesiologists practicing in metropolitan areas was significantly greater in the monitoring group relative to the non-monitoring group ($p = 0.026$).

Statistical comparison of Kaplan-Meier curves, using the log-rank test, indicated that the anoxic oxygen supply monitoring subgroup made correct diagnoses significantly sooner than the non-monitoring subgroup ($p = 0.019$). The pneumothorax monitoring subgroup also tended to make correct diagnoses sooner than the non-monitoring subgroup, although this difference was not significant ($p = 0.066$). Cox regression hazard ratios, all greater than 1, indicated that monitoring subgroups were more likely to make correct diagnoses than non-monitoring subgroups within any of the critical incident groups, although 95% confidence bounds for the hazard ratios all bracketed 1 (i.e., not significant). Fisher

exact probability test comparison of the proportion of anaesthetists making correct diagnoses within 8 minutes did not differ significantly between the monitoring and non-monitoring subgroups in any of the critical incident groups, although the monitoring subgroup proportions were always greater than or equal to non-monitoring subgroup proportions.

Table 2 shows the times to diagnosis for the four different scenarios run with and without pulse oximetry and capnography. The two scenarios with the greatest incidence of "no diagnosis after 8 minutes" (pulmonary embolism and anoxic oxygen supply, both without capnography and pulse oximetry) also had the highest incidence of drug use. Repeated administration of a given medication counted only once.

Assuming exponentially distributed diagnosis times, our power analysis predicts that a simulator-based study with 8-minute simulation runs would require a total of 3,097 (or 438 or 87) equally allocated subjects to have an 80% chance of establishing a minimum 10% (or 25% or 50% respectively) reduction in mean time to diagnosis relative to a mean control time of 6 minutes at a significance level of 0.05. Our power analysis also predicts that for longer simulation runs of 10 and 12 minutes respectively, the sample size would fall from 3,097 to 2,800 and 2,618 respectively.

DISCUSSION

Simulators enable the presentation of clinical scenarios that are rare or could not be ethically performed on human patients. For example, in a single blind study of pulse oximetry in children, Cote et al. [10] decided to let the anesthesiologists know when SpO_2 values dropped to below 85% for more than 30 s. To investigate severe conditions in patients it is necessary to wait until such conditions present unexpectedly. Consequently, large studies involving a large patient population are needed to test hypotheses regarding rare events [1–3, 11, 12].

A patient simulator enables the presentation of identical critical incidents to many clinicians. This eliminates the many confounding differences that exist among patients and treatments. The reproducibility of the patient simulator addresses directly the requirement for comparable cases stated by Duncan & Cohen [4] and specified by Cook et al. [13] that "experimental designs should include critical incidents from comparable monitored and unmonitored situations." Simulation also provides an answer to the challenge issued by Cook et al. [13] to "devise practical methods of uncensored and detailed investigation of incidents as they occur." By eliminating confounding influences in the "patient" population and using an efficient analytic tool such as the Cox proportional hazards model, the sample size requirement can be greatly reduced. As an example, our power analysis prediction of 3,097 equally allocated subjects in a simulator-based study compares favorably with the estimate of "7,546,605 cases… to have a reasonable chance (0.80) of achieving a statistically significant difference ($p < 0.05$)" in patient outcome as a result of adoption of the Harvard monitoring standards [12, 14]. The reduction in sample size should translate into a significant cost reduction of conducting a study to test the effectiveness of existing or prototype medical devices. It is conceivable that in the future, the efficacy of new medical technology will be prospectively evaluated using patient simulator-based studies.

The issue of the temporal relationship of the introduction of the monitor relative to the time period during which the study is conducted has been raised by Duncan & Cohen [4]. An ideal study would be performed with a clearly delineated "time zero." Before "time zero," the new monitor would ideally never have been used and after "time zero" all cases would be using the new monitor, for a comparative outcome study. In reality, it is hard to pinpoint "time zero" as monitors are usually introduced gradually. With the simulator-based study we designed, there is no need to establish "time zero" because the study occurs in a compressed time frame (5 days) and the design of the study focuses on the presence or absence of the monitor rather than the time of introduction of the monitor.

It might seem contradictory that we reached statistical significance for the anoxic oxygen supply scenario using only 22 subjects when our power analysis predicted that we needed 3,097 subjects. It should be noted that the log rank test which was used to establish significance for the anoxic oxygen supply incident is a non-parametric test. The power analysis, on the other hand, is based on a parametric test, which assumes that the underlying behavior of the population is known (in our case, we assumed a constant hazard function and exponentially distributed diagnosis times). The power analysis also accounts for the *magnitude* of the effect. For our power analysis, we assumed that a 10% reduction in mean diagnosis time was clinically relevant. As the percentage difference, or the magnitude of the change becomes smaller or subtler, the sample size required to establish significance increases. Conversely, if the effect is gross, a smaller sample size is required.

The power analysis predicts that simulation run time also affects the sample size. Longer simulation runs reduce the sample size because more subjects can reach a diagnosis in the additional time allowed. However, the simulation run time has to be balanced against the

Table 2. The time in seconds to diagnosis of the critical incidents

	PTX: Yes	PTX: No	PE: Yes	PE: No	MH: Yes	MH: No	AOS: Yes	AOS: No
	60	96	216	180	120	144	312	432
	96	108	252	240	132	180	324	480
	108	156	300	264	144	204	348	480
	120	204	396	288	180	216	396	480
	120	204	408	480	180	288	432	480
	120	216	432	480	264	312	432	480
	132	228	432	480	300	372	480	480
	156	264	444	480	300	420	480	480
	168	276	480	480	432	420	480	480
	180	336	480	480	468	444	480	480
	216	480	480	480	480	468	480	480
	432			480		480		
Average	159.00	233.45	392.73	401.00	272.73	329.00	422.18	475.64
Standard deviation	95.33	108.05	94.04	119.17	135.99	120.81	66.99	14.47
Median	126	216	432	480	264	342	432	480

Abbreviations: PTX – pneumothorax, PE – pulmonary embolism, MH – malignant hyperthermia, AOS – anoxic oxygen supply.
Yes – with SpO_2 and CO_2 data; No – without SpO_2 and CO_2 data. Critical incidents not diagnosed within 8 minutes were assigned a score of 480 seconds. A table documenting the frequency with which the participants used medications in the management of the different scenarios can be viewed at http://gasnet.med.yale.edu/periodical/jcmc/1998/July as well as two tables depicting typical anoxic oxygen supply cases, with pulse oximetry and capnography data visible to the participant.

clinical reasonableness (unrealistic to have a long simulation run for an incident that develops and concludes rapidly) as well as the logistics of the study.

In designing and conducting the experiment we had to accept a number of compromises. The setting differed from a real operating room in several respects, not the least of which were the noises intruding from the exhibition hall. Even though no participant complained, we assume that it presented a distraction. All participants were also aware that a critical incident was likely to occur during the time they were caring for the patient but this potential influence applied to both groups. We recognize the possibility that some participants on the first or second day of the exercise reported their experience to participants of day 3 and 4. However, the data do not show better performance on the later days of the study. The average and standard deviation of the time to diagnosis was 333 ± 120 s for day 1, 333 ± 135 s for day 2, 329 ± 144 s for day 3 and 342 ± 163 s for day 4. Further, the subject was not alone with the instructor; there were also a cameraman and a simulator technician in the room at all times, which may have induced performance anxiety or stage fright in some subjects.

We kept the scenarios relatively short so that we could collect data from about 100 subjects during the 4-day meeting. While a short scenario was quite sufficient for the anoxic oxygen supply incident, it was too short for the malignant hyperthermia exercise and adequate but not ideal for the pneumothorax and the pulmonary embolism drills. Yet, none of the clinicians commented on the faster than real time aspect of the simulations. It can also be debated whether 2 minutes was enough for the subject to become familiar with the environment before triggering the incident.

The assumption of a constant hazard function resulting in exponentially distributed diagnosis times implies that the instantaneous probability of making the correct diagnosis is constant throughout the simulation run. It can be argued that through a process of elimination of incorrect diagnoses as time progresses, the probability of a correct diagnosis increases and should not stay constant. Characterization of the increased probability of a correct diagnosis as time elapses needs to be further studied.

The patient simulator we used internally measures the alveolar concentration of inhalational agents and automatically responds accordingly. However, because the exhibition hall provided no scavenging system, we could not use this feature of the simulator and had to simulate an anesthetic without recourse to halogenated inhalation anesthetics. This logistic constraint explains the fairly high dose of morphine that we used in the case stem, which elicited comments from several participants.

The scenarios we selected were chosen because all of them would exhibit certain clinical signs, which, together with historical clues, would give the clinician a chance to arrive at the correct diagnosis without pulse oximetry and capnography.

Some clinical signs of current full-scale simulators fall short of realism. The clinical signs are adequate to discern changes in compliance and breath sounds but inadequate by failing to present changes in skin temperature and color.

Data that would normally have been obtained from the surgical field was provided at the instructor's discretion or when the subject inquired. Potential inconsistency in providing the surgical data among the instructors, who were not blinded to the scenario, could have biased the study. In hindsight, it would have been preferable to provide identical data from the surgical field at consistent, scripted times for all scenarios, just like the gas cylinder was changed for all scenarios. Such a change in the protocol would also have allowed us to blind the instructor to the scenario to avoid introducing bias. During our study, the instructors were not blinded to the critical incident because they needed to know what clues to provide from the surgical field.

We started the stopwatch for measuring the time to diagnosis at the first observable change in different vital signs depending on which critical incident was being simulated. If comparison of diagnosis times *between* critical incidents is a study objective, future researchers should consider using the first observable change in a standardized vital sign, e.g., heart rate, to start the timer, for all critical incidents.

Response patterns to the challenges of the scenarios varied greatly among the participants. Many participants switched from mechanical to manual ventilation at the first sign of trouble even though a pressure gauge on the anesthesia machine displayed the peak inspiratory pressure and a respirometer as well as the ventilator bellows provided information on the tidal volumes delivered. We do not know how much of this response is attributable to the desire to have moment to moment information on compliance and tidal volume and how much is an expression of the fiddle response. On the other hand, manual ventilation and the resulting uneven ventilation may vitiate the interpretation of end-tidal PCO_2, particularly when capnographic data are important in the diagnosis. One participant misinterpreted the low end-tidal carbon dioxide (owing to manual hyperventilation) as attributable to inadequate pulmonary blood flow.

None of the scenarios called for the pharmacologic treatment of the condition and none of the conditions were amenable to correction by a drug – with the exception of malignant hyperthermia where we stopped the scenario as soon as the diagnosis had been made. A number of drugs were used in support of the circulation; others had no obvious indication. The use of drugs in these circumstances is an interesting area for future studies.

Clinically the most difficult scenario was the disappearance of oxygen from the gas supply to the anesthesia machine. When participants were told that the oxygen analyzer was broken, most proceeded with the case as if the oxygen supply was intact. Participants who switched to the self-inflating bag often asked for supplementary oxygen – without realizing that it might also contain a gas other than oxygen. Although oxygen analyzer failure may at first appear contrived, death by hypoxemia due to failure of oxygen delivery continues to occur worldwide, even with widespread availability of O_2 analyzers. O_2 analyzers fail, especially if the sensor cells are not replaced in a timely fashion. Bad outcomes rarely result from a single failure; they generally result from two or more failures. Because one of our objectives was to study the influence of pulse oximetry and capnography on the time to diagnosis of an anoxic oxygen supply, we did not want the F_IO_2 sensor to provide an early warning of problems with the oxygen supply.

During the anoxic oxygen supply exercise, switching from ventilator to bag and manual hyperventilation was often associated with a temporary improvement, probably secondary to the fact that the breathing bag still contained oxygen, having been in the breathing circuit before the misfilled cylinder had been attached.

The present study, despite its limitations, points to the potential value of simulation in the assessment of diagnostic and monitoring information. The fact that the differences between the groups were smaller than anticipated points to the importance of history and physical findings, even in simulated exercises. It supports the observation that modern monitors supplement rather than supplant clinical skills.

We gratefully acknowledge support from Medical Education Technologies, Inc., Paterson Distributors Ltd., Ohmeda, Inc., SpaceLabs, the United States Anesthesia Patient Safety Foundation and the organizing committee of the 11th World Congress of Anaesthesiologists. Special thanks to our simulator instructors from Australia and New Zealand (Steuart Henderson, Brian Lewer, Jennifer Weller, Leona Wilson (Wellington Hospital); William Runciman, John Russell, Robert Webb, John Williamson (Royal Adelaide Hospital); Stuart Montano, Anne See (Royal North Shore Hospital); John Gibbs (Dunedin Hospital); Richard Horton (Austin Hospital); Brian Robinson (Medical College of Wisconsin), Ross Paterson and Simon Wilson (Paterson Distributors), Linda Sielaff (Ohmeda, Inc.) and Tom Crane, Paul McGee, Cyndy Burkhardt, Wayne McAuliffe and George Chrys (SpaceLabs Medical Products Pty, Ltd.). Finally, we thank John Hardcastle, Hans van Oostrom, Ron Carovano, Michael Ellis, Jan Beneken, Kelly Spaulding, Fame Cricks and Ryan Chin for their invaluable assistance.

REFERENCES

1. Moller JT, Pedersen T, Rasmussen LS, Jensen PF, Pedersen BD, Ravlo O, Rasmussen NH, Espersen K, Johannessen NW, Cooper JB, Gravenstein JS, Chraemmer-Jorgensen B, Wiberg-Jorgensen F, Djernes M, Heslet L, Johansen SH. Randomized evaluation of pulse oximetry in 20,802 patients. I. Design, demography, pulse oximetry failure rate, and overall complication rate. Anesthesiology 1993; 78: 436–444
2. Moller JT, Johannessen NW, Espersen K, Ravlo O, Pedersen BD, Jensen PF, Rasmussen NH, Rasmussen LS, Pedersen T, Cooper JB, Gravenstein JS, Chraemmer-Jorgensen B, Djernes M, Wiberg-Jorgensen F, Heslet L, Johansen SH. Randomized evaluation of pulse oximetry in 20,802 patients. II. Perioperative events and postoperative complications. Anesthesiology 1993; 78: 445–453
3. Moller JT, Svennild I, Johannessen NW, Jensen PF, Espersen K, Gravenstein JS, Cooper JB, Djernes M, Johansen SH. Perioperative monitoring with pulse oximetry and later postoperative cognitive dysfunction. Br J Anaesth 1993; 71: 340–347
4. Duncan PG, Cohen MM. Pulse oximetry and capnography in anaesthetic practice: an epidemiological appraisal. Can J Anaesth, 1991; 38 (5): 619–625
5. Lampotang S, Good ML, Westhorpe R, Hardcastle J, Carovano RG. Logistics of conducting a large number of individual sessions with a full-scale patient simulator at a scientific meeting. J Clin Monit 1997; 13: 399–407
6. Lampotang S, Good ML, van Meurs WL, Carovano RG, Azukas J, Rueger EM, Gravenstein JS. The University of Florida/Loral human patient simulator, abstracted. J Anesthesia 1995; 9: SS1–5
7. Good ML, Lampotang S, Ritchie G, Heffels J, Miller B. Hybrid lung model for use in anesthesia research and education, abstracted. Anesthesiology 1989; 71: A982
8. Van Meurs WL, Good ML, Lampotang S. Functional anatomy of full-scale patient simulators. J Clin Monit 1997; 13: 317–324
9. Rosner B. Fundamentals of Biostatistics, 4th ed. Belmont, CA: Duxbury Press, 1995
10. Cote CJ, Goldstein EA, Cote MA, Hoaglin DC, Ryan JF. A single blind study of pulse oximetry in children. Anesthesiology 1988; 68: 184–188
11. McKay WPS, Noble WH. Critical incidents detected by pulse oximetry during anesthesia. Can J Anaesth 1988; 35 (3): 265–269
12. Eichhorn JH. Prevention of intraoperative anesthesia accidents and related severe injury through safety monitoring. Anesthesiology 1989; 70: 572–577
13. Cook RI, Woods DD, McDonald JS. On attributing critical incidents to factors in the environment (letter to the editor). Anesthesiology 1989; 71: 808
14. From RP, Pearson K, Tinker JH. Did monitoring standards influence outcome? (letter to the editor). Anesthesiology 1989; 71: 808–809