Standardized Simulated Events for Provocative Testing of Medical Care System Rescue Capabilities

George Blike, Joseph Cravero, Steve Andeweg,
Jens Jensen, Klaus Christoffersen

Abstract

Background: Human errors that result in rare critical events during the course of routine medical care are especially difficult to study. Although the incidence of severe respiratory depression with routine sedation is small, millions of patients receive this care annually, putting hundreds or thousands of persons at risk. Elimination of sedation errors is particularly important since associated deaths and neurological injuries are virtually 100 percent avoidable. **Objectives:** The researchers tested the feasibility and validity of using a commercial human simulator to probe care systems for system vulnerabilities ("accidents waiting to happen"). **Methods:** The study required (1) development and validation of a simulated rare event, (2) use of the standardized event to test care systems in context, and (3) video analysis for deviations in observed care relative to gold-standard care. *Scenario*: The standard event was reproducible, with physiology that degraded over time if no interventions occurred, and improved when treated appropriately. *"Crash-testing" actual care domains:* Team performance in managing the simulated event was videotaped and data files of the simulator's physiology variables were captured. *Analysis:* The quality of each team's performance was assessed using the simulator data files to calculate "time-out-of-range" measures for the critical variables, and team behaviors were analyzed for deviations from idealized care. **Results:** Available technology supports the creation of a sedation critical event that is realistic, reproducible, and portable. The simulator-based provocative test readily allowed comparison of rescue performance in different sedation care settings (e.g., emergency room, radiology department) to be contrasted with a gold standard. **Conclusion:** This research supports the feasibility of using available human simulation as a "crash-test dummy," capable of measuring the rescue systems used in a variety of actual sedation care settings. The findings demonstrate that personnel deemed competent and safe, on the basis of meeting hospital training requirements for airway management, had profound performance deviations when compared to gold-standard practice.

Introduction

Human error and a lack of patient safety are recognized as leading causes of preventable death in America.[1] As we have noted in previous studies,[2] managing the pain and anxiety associated with medical procedures is complex and presents

typical patient safety challenges. The pressure to satisfy patient desires for painless, stress-free care for all types of procedures has led to an explosion in the use of more potent sedative medications by nonanesthesiologists. For example, the use of ketamine and propofol by gastroenterologists, emergency medicine physicians, and intensivists has grown significantly in the past several years.[3–5] The safety of these practice changes have proven difficult, if not impossible, to assess with prospective studies due to the number of patients that would be required.

Although severe respiratory depression is relatively rare, millions of patients are exposed to this risk, resulting in an aggregate morbidity and mortality that is noteworthy. Large, sufficiently powered, multicenter trials to evaluate safety in light of these changing practices simply do not exist. Instead, the literature is replete with descriptions of how sedative medications can be used in a variety of settings on a series of patients (usually fewer than 200 in a cohort) without a fatality.[6–8] Given the expected incidence of an induced crisis with routine sedation, it is not surprising that these studies rarely uncover a critical event. The authors invariably conclude that the described sedation practice is safe.[9] In the absence of critical incidents however, safety cannot be presumed since no data are available to assess the efficacy of rescue processes.

Sedation care delivery systems that lack the ability to manage respiratory depression pose a serious threat to patients. In the mid-1980s, 86 deaths in the United States were attributed to the just-released anxiolytic, midazolam.[10] Despite the introduction of guidelines for safe sedation care and Joint Commission on Accreditation of Healthcare Organizations (JCAHO) standards, "conscious sedation" for minor procedures remains associated with an alarming number of fatalities. Cote et al.[11] published a critical incident review of 90 pediatric sedation-related critical incidents resulting in death or severe neurological injury. The critical incident analysis determined that the overwhelming majority of these deaths were avoidable and associated with a "failure to rescue" and a failure to provide airway and ventilatory support in a timely fashion. This report has served as a warning that pediatric patients are at especially high risk for injury during sedation.

Our research investigated the feasibility and potential impact of a new methodological approach for measuring patient safety in health care. Pediatric sedation (i.e., the acute treatment of procedure-related pain and stress with medications in children) served as the test domain for this research. We explored the potential to use available human simulation to provocatively test current sedation care settings for safety and rescue capability in an objective manner. A pediatric simulator was programmed to create a realistic sedation event that posed multiple challenges and required rescue interventions. Rescue performance that defined a gold standard was generated by a pediatric anesthesiologist who performs a high volume of pediatric sedations, using standard tools and techniques. The same simulated event was used to "crash test" actual care settings where sedation care is provided. Finally, we used quantitative and qualitative methods of analysis to review videotapes and simulator physiologic data files

collected during the exercises to identify care management problems. In summary, this research seeks to test the feasibility and validity of using a human simulator to probe care systems for latent system failures (system vulnerabilities or "accidents waiting to happen").

Methods

Test domain

Pediatric sedation for the acute treatment of procedure-related pain and stress was used as the test domain for this investigation. This area of treatment was chosen because children are a high-risk, low-error-tolerance subset of all patients receiving sedative medications. The smaller anatomy, increased rate of desaturation, and more difficult airway management causes children (especially neonates) to be more likely to experience errors in management, and those errors are more likely to result in negative outcomes.

Critical event selection and refinement

Based on the critical incident review of pediatric sedation-related deaths, respiratory depression leading to airway obstruction and/or central apnea was the most serious event associated with a negative outcome when rescue systems failed.[11] We chose to design a single scenario that presented a child receiving sedation who develops an obstructed airway, followed by frank apnea. If treatment was properly instituted for airway obstruction with an oral airway and a chin lift, simulated obstruction was reversed and the patient was left only with apnea. If bag-mask ventilation was provided properly, the hypoxia would resolve. If providers did not have proper equipment available or failed to use resuscitative equipment appropriately, the simulated child would remain hypoxic and exhibit the hemodynamic consequences of hypoxia over time.

Development and validation of the simulated critical sedation event

The simulator platform used was that of the METI™ (Medical Education Technologies, Inc., Sarasota, FL) pediatric human simulator. The design and validation strategy consisted of (1) defining the expected behavior for the natural evolution of apnea in a child associated with sedative medication overdose, (2) testing of observed simulator behavior, and (3) making software and hardware modifications to the simulation until the observed behavior and the modeled behaviors matched.

Expected behaviors. Three physicians who practice pediatric anesthesia, pediatric intensive care, and pediatric cardiac anesthesia participated in a structured interview to define the onset and cessation of hypoxia, and hypoxic cardiac response associated with drug-induced apnea. Empiric data on typical rates of desaturation (based on lung volumes and functional residual capacity in a

normal 5-year-old child), along with expected respiratory mechanics were derived from the available literature.*

Observed simulated event behavior. While an apneic event was initiated on room air and 100 percent O_2, the PaO_2, SaO_2, heart rate (HR), and blood pressure (BP) were captured in a data file. Respiratory mechanics were directly measured using an intensive care unit-(ICU-) grade ventilator (Siemens 2000; Siemens Medical Solutions USA, Inc., Malvern, PA).

Hardware and software modifications. An iterative process was used to align expected behavior with that of the observed behavior. Three levels of complexity for interventions were created, such that success with level 1 would trigger the next level of patient complexity and require more advanced rescue interventions. This approach would allow sedation system performance to be challenged at both low and high levels of task complexity.†

Gold-standard performance. To provide a reference against which to compare the performance of sedation care in the various settings in which it is provided, we assessed rescue performance in managing the simulated scenario and embedded events under "ideal" conditions. An experienced pediatric anesthesiologist (JC), who was not given knowledge of the scenario features in advance, provided sedation and managed the scenario and apneic events using a standard complement of resources that represented the standard-of-care provided by anesthesiologists working in the pediatric procedural sedation unit of our children's hospital. Such resources have been in place for more than 2 years, governing approximately 1,300 sedations per year. This served to define a demonstrably attainable standard of performance, both in terms of the process followed and in terms of how well the critical variables were maintained within their desired ranges.

Test settings. Several settings were selected to test the method. Pilot studies were conducted in the interventional radiology department and the emergency medicine department. Both settings provide pediatric sedation care on a regular basis. When a "code blue" is initiated at Dartmouth Hitchcock Medical Center (DHMC), responders include RNs; flight nurses; critical care nurses from adult, pediatric, and neonatal ICUs; respiratory therapists; anesthesiology residents; physician supervisors of the sedation provider; secretarial support; and administrative support. Code responders were not alerted to the fact the response was a simulation exercise until arrival and direct observation of the simulated patient. The validated scenario was used to test the rescue performance of the Pain-free Pediatric Sedation Unit at the Children's Hospital at Dartmouth (CHaD).

Data collection. Data were obtained from video recordings of each team's performance, and from the simulator output log describing the physiological state

* Data from pediatric anesthesia experts and parameters used to depict pulmonary function for a 5-year-old, 20 kg child, 112 cm tall, is available at: http://an.hitchcock.org/ahrq/Blike2005.pdf.
† The scenario code is available at: http://an.hitchcock.org/ahrq/Blike2005.pdf.

of the simulated patient. The video recordings, done with a Sony Digital 8 HandyCam®, captured a single wide-angle view of rescue performance by the care team and code responders. The video was recorded from the foot of the patient bed, looking down from a height of 8 feet. Monitors, equipment, clinicians, interventions, and the simulated patient were all visible on the tape. The camera was panned to cover the entire range of activities within the confines of the room in which care was being provided. Audio was captured using the single microphone integrated in the digital video camera.

Quantitative analysis. The simulator's output was used to perform quantitative analyses of the quality of control exercised by the teams over the patient's physiology. Video markers had been developed that represented treatment milestones. Phase I care was based on first responder performance of monitoring and detecting the event, mobilizing rescue resources, and performing basic airway maneuvers of verbal stimulation, jaw lift, and use of an oral airway. Phase II care consisted of advanced airway management with the arrival of providers who are experts in the provision of positive pressure ventilation, the placement of a definitive airway, and the provision of Pediatric Advanced Life Support. Hypoxia and hypotension were defined as SpO_2<60 percent and systolic BP<50 mm Hg, respectively, as these parameters would be associated with negative patient outcomes over time. Time out-of-range for each of these parameters was calculated for the gold-standard exercise, and the exercises within actual sedation care settings. This method of analysis has been previously described for comparison of simulator performance between novices and experts.[12]

Qualitative analysis. The video recordings were analyzed to produce process traces that described the behavior of each team during the case, and to identify specific care management problems.[13] These problems were further analyzed to identify contributing factors using a published taxonomy.[14, 15] The goals of this analysis were to identify deviations in performance from the standard-of-care that was proven possible in the gold-standard performance exercise. For each care management problem observed by experts, contributory factors were listed that represented the latent error in the system of care (i.e., the accidents that were waiting to happen under the right triggering conditions). This methodology is currently used by the Dartmouth Hitchcock Department of Anesthesia Quality Assurance Committee to review over 200 critical incidents and/or close call reports.[16] All three expert reviewers (pediatric anesthesiologists practicing at CHaD) had experience performing critical incident review using the methodology outlined by Vincent et al.[15] In addition, all three experts reviewed the video-record of the gold-standard case prior to reviewing the video-record of care in the actual sedation care setting. Contrasting observed care against "ideal care" was intended to reduce reviewer variability.[17] Individual review data were compiled to fill a data table template (Table 1A). The aggregate data were reviewed by the group to confirm face validity. If two of three reviewers felt an identified performance deviation was trivial and unlikely to have the potential for patient harm, it was deleted.

Table 1. Video markers of gold standard and test sedation

1A. Monitors and rescue equipment

Video-markers of "ideal" monitors and rescue equipment			Gold standard	Test sedation unit
Monitoring	Ventilation	Direct Observation	1	1
		$ETCO_2$	1	0
		Cont. auscultation	1	0
	Oxygenation	SpO_2	1	1
		Cont. Tone/Beep	1	1
		Alarm for SpO_2	1	1
	Perfusion	SpO_2 pleth	1	1
		SpO_2 HR	1	1
		EKG HR	1	0
		EKG trace	1	0
		NIBP	1	1
	Level of Consc.	Verbal	1	0
		Pain	1	0
Deviations		13 markers	0	6
Rescue Equipment	For regurgitation	suction/yankauer	1	1
	For PPV	Bag	1	1
		Mask	1	1
		O_2 source	1	1
		Oral airway	1	1
		LMA	1	0
	Definitive airway	Laryngoscope	1	1
		ETT	1	1
		stylet	1	1
	Reversal drugs	Narcan	1	1
		Flumazenil	1	0
	Call for Help	Back-up available	1	0
		access mechanisim	1	1
Deviations		13 markers	0	3

The experts developed a tool for itemizing the equipment and behaviors associated with best practice and time ranges that they agreed were reasonable for categorizing performance as good, intermediate, or poor (Table 1B).

Results

Development and validation of the simulated critical sedation event

Observed simulated apneic event behavior. The METI pediatric simulator was programmed to be a 5-year-old child of the height and weight specified and, after a period of stability, apnea was initiated with an instantaneous 100 percent neuromuscular blockade. The simulator needed to be programmed to have abnormally high oxygen consumption to develop hypoxia, and it was therefore set

1B. Obstructive and central apnea "behaviors"

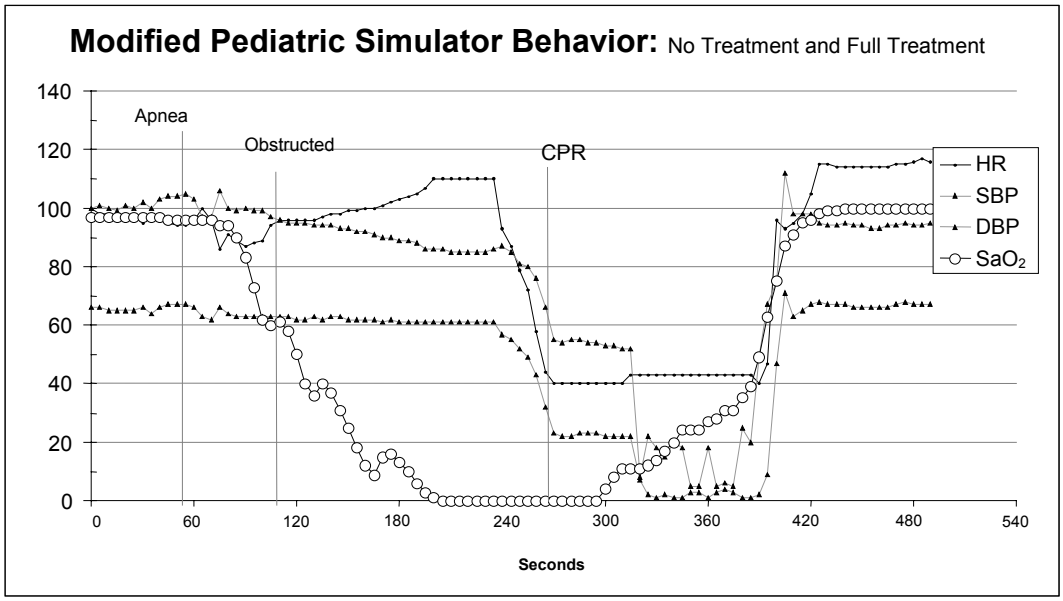
Video-markers of "ideal" obstructive and central apnea "behaviors"			Gold standard	Test sedation unit	Scoring criteria (seconds)		
					Good	Adequate	Poor
Phase 1	Monitoring	Apnea diagnosed (no chest movement)	18	38	0–30	31–60	>60
	Mobilizing help	PPV call (from time apnea detected)	0	11	0–30	31–60	>60
	Basic Airway Tx	Supplemental O_2 (from time apnea detected)	3	0	0–15	16–60	>60
		Jaw Lift (from time apnea detected)	5	15	0–15	16–30	>30
		Oral Airway (from time jaw lift)	15	52	0–15	16–30	>30
		Bag/Mask Ready (from time requested)	3	13	0–15	16–30	>30
Aggregate time to complete Phase 1 tasks			**44**	**129**			
Phase II	Advanced Airway Tx	PPV expert arrives (from time called)	0	182	0–120	120–240	>240
		Expert BMV (PPV attempts from when arrived)	7	>30	0–15	16–30	>30
		Two Person (from when one person failed)	0	>30	0–15	16–30	>30
		Intubation (from when two person failed)	NA	115	0–60	61–120	>120
		Failed intubation "call for back-up" (from when failed)	NA	NA	0–15	16–30	>30
		Succinyl Choline (from when laryngospasm dx)	NA	NA	0–60	61–120	>120
	PALS	Atropine (HR<60)	NA	NA	0–60	61–120	>120
		Epinephrine (Atropine, HR<60)	NA	NA	0–60	61–120	>120
		Compressions (no pulse)	NA	NA	0–15	16–60	>60
Aggregate time to complete Phase II tasks			**16**	**297**			
Time out of range: SpO_2< 60%			**0**	**90**			

at the maximum of 2,000 ml per minute. Internal physiological data files captured the development of hypoxia and hemodynamic behavior over time. The same experiment was conducted after having the simulator intubated and on 100 percent oxygen for 5 minutes prior to initiation of the apneic event. In both instances the simulator did not behave in a realistic manner, taking much longer to desaturate than expected.

Hardware and software changes. The current model of the pediatric simulator only used CO_2 insufflation to create hypoxia at the mechanical sensor. Adding a T-connection to the CO_2 inlet line for a nitrogen line that included a calibrated nitrogen flow regulator allowed creation of more realistic behavior. In addition, a scenario was created in the software to better model the expected hemodynamic response to hypoxia described by our experts.‡

The measured behavior of the simulator after hardware and software modifications approximated the expected behavior in terms of the onset of hypoxia as a function of the FiO_2 at the time of apnea, and the hemodynamic response to both untreated hypoxia and the response to rescue (Figure 1). The standard event was reproducible, with physiology that degraded over time if no interventions occurred, and improved when treated appropriately (Figure 2). Performance in managing the simulated event was videotaped and data files of the simulator's physiology were captured. The data recordings were graphed and annotated (Figure 3).

Figure 1. Behavior of the modified pediatric simulator



Legend: Experiment in which pediatric simulation (with hardware and software modifications) was allowed to progress after initiation of the apneic event with no therapeutic interventions, followed by full resuscitation using Pediatric Advanced Life Support algorithms.

‡ A detailed description of these modifications is available at: http://an.hitchcock.org/ahrq/Blike2005.pdf.

Figure 2. Physiological data of simulated sedation scenario managed by experts (ideal care setting)

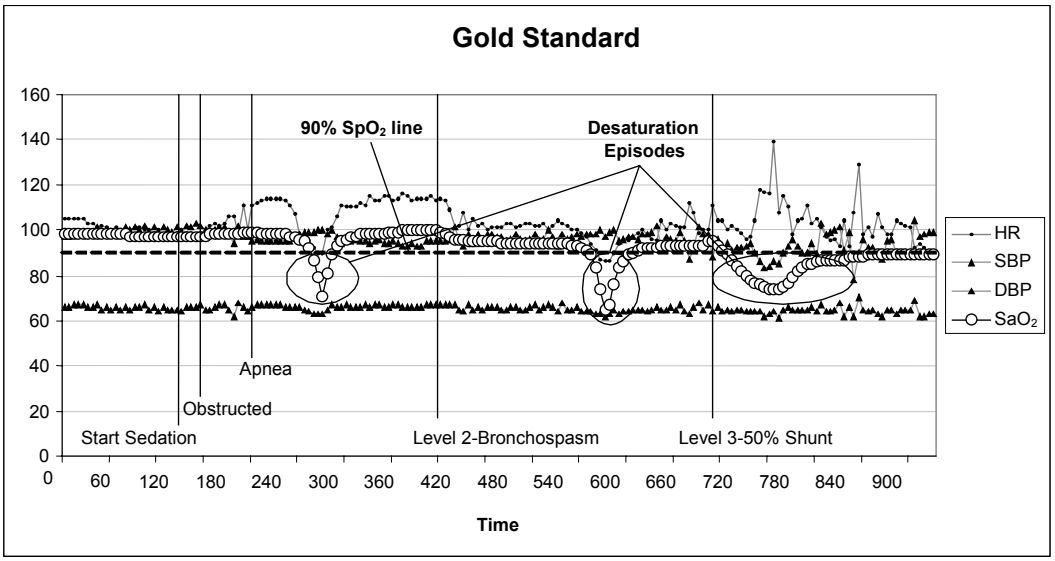
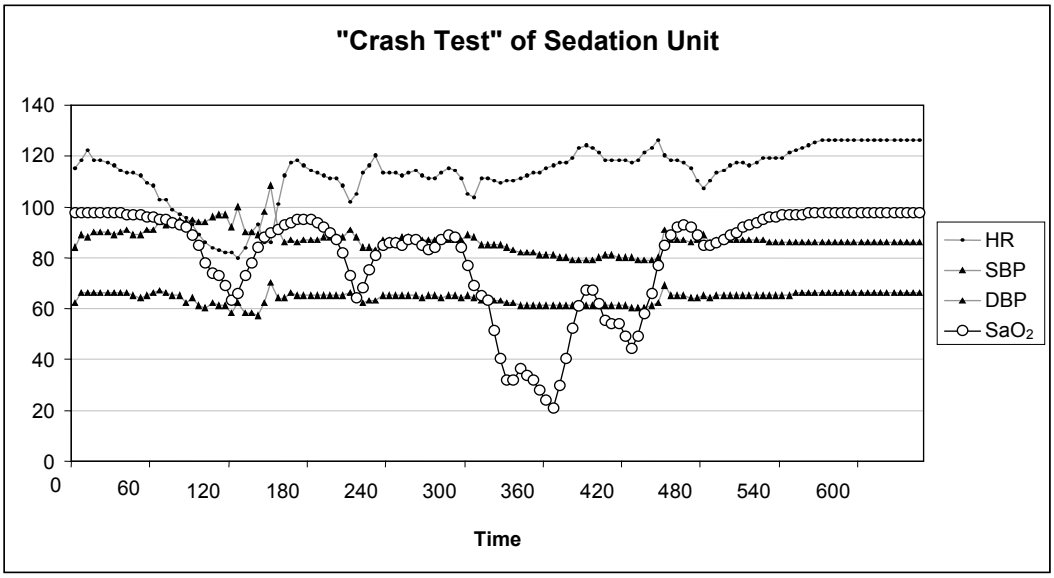


Figure 3. Simulated sedation emergency (nonideal conditions)



Video and physiological data analysis. Rescue equipment, interventions, and critical task execution were captured on video (Table 1A), which was scored for the initiation and completion times for a subset of these tasks from the onset of obstruction/apnea. Time-out-of-range was then calculated for the critical variables (Table 1B).

Compared to ideal, the sedation unit tested had six deviations in monitoring and three deviations in using rescue equipment. Treatment milestones failed to be met in two of the Phase I tasks (the diagnosis of apnea and use of an oral airway)

and four of the Phase II tasks (arrival of the expert airway provider, the expert performing positive pressure ventilation or attempting two-person ventilation, and the time to achieve intubation). Total times to complete tasks were 85 seconds longer (for Phase I) and 281 seconds longer (for Phase II) in the actual-care setting when compared to the gold standard. Both settings avoided the terminal bradycardia and hypotension that will develop if hypoxia is severe and prolonged. Total time with severe hypoxia (SpO_2<60 percent) in the actual care setting was 90 seconds. The scenario did not progress to Level 2 and Level 3 difficulty, due to intubation being performed during the Level 1 event.

Qualitative analysis. Video review by pediatric anesthesia experts identified care management problems and the contributory factors associated with each (Table 2).

Discussion

Identifying hazards and vulnerabilities in the performance of complex health care delivery processes is a methodologically challenging problem. Error reporting systems provide limited detail and are subject to various reporting biases.[18] Direct observation has proved useful for gaining detailed insight regarding performance of complex systems. However, techniques that rely on opportunistic observation of actual cases are ill-suited for the investigation of relatively rare events (e.g., respiratory arrests due to oversedation in pediatric patients).[19]

We sought to develop a reproducible method that could be used to assess the systems that provide sedation care in a typical hospital. We used an interactive patient simulator to model a classic oversedation response for a pediatric patient. Instead of bringing care teams into a simulator center as is often done, we designed the scenario as an event running on a portable system that could be brought into actual sedation care settings to observe how the teams in these settings responded to manage the event. This corresponds to what human factors researchers refer to as a "field experiment," in which certain aspects of a naturally occurring work situation are deliberately manipulated by investigators to permit targeted observations of a specific type of problem-solving scenario.[13] This class of investigations is intended to allow observation of behavior under conditions that are highly representative of actual work conditions[§].[20] Specifically, we wanted to create conditions that would allow us to observe performance using the actual resources (i.e., the personnel, equipment, and procedures) that would be brought to bear on a real patient experiencing the same type of event. In this study, we extended previous work that used a gold-standard comparison with both quantitative and qualitative forms of analysis.[3]

Our primary conclusion is that this methodological approach is technically feasible and generates powerful data regarding system performance during the

[§] We prefer this more precise term to the commonly used concept of "ecological validity" (cf. Hammond, 1998).

Table 2. Qualitative review to identify rescue system problems in an actual sedation unit

Care Management Problems	Description	Contributory Factors
1. Event Detection Errors A. Verbal stimulation for Level of Sedation not used B. Delay in diagnosis of apnea in setting of desaturation	RN appropriately reviewed history, identified SpO_2 as decreased, and provided face mask oxygen. However, she did not assess for responsiveness to verbal stimulation, provide painful stimulation via jaw lift, or visually assess respiration (chest rise). Instead she used auscultation and took >30 seconds and then did jaw lift.	**Patient**-Cerebral Palsy and Seizures **Task**-none contributory **Practitioner**-RN did not assess airway and ventilation immediately upon identification of desaturation episode **Team**-no **Working cond**-Phase 2 recovery location, nursing ratio 1:2 **Organization**-Sedation training not standardized **Political**-no
2. Initial Management Errors A. Did not stimulate B. Oral airway selection delay C. Technique in BM ventilation suboptimal, 2 person not attempted D. Cause of apnea, potential reversal with naloxone not considered	RN after identifying apnea, requested help from anesthesiologist and attempted ventilation. They did not stimulate verbally or aggressive jaw lift. They did not insert oral airway immediately with supplemental oxygen. They then requested a "code blue" be initiated as the oxygen saturation worsened. Bag-mask ventilation was initiated with hand position such that mechanical leverage for opening airway was suboptimal. Did not request help with positive pressure ventilation. Naloxone not considered.	**Patient**-Airway obstruction, then apnea **Task**-Airway management "safing" practices not standardized **Practitioner**-PCT unable to select correct size oral airway **Team**-RN not as directive as could be. When resident members of "code team" arrived, did not contribute **Working cond**-RN covering recover from sedation is on their own for initial resuscitation **Organization**-Sedation training not standardized **Political**-no
3. Diagnostic Decision-making A. Alternate hypothesis, broad differential diagnosis not stated.	RN did not act as "team leader" in the sense of declaring the situation, assessment, and actions that need to be taken. When code team members arrived, they also failed to establish leadership. Differential diagnosis never stated aloud, naloxone never considered, pauses in bag-mask ventilation not identified or managed while anesthesiologist setting up to intubate.	**Patient**-Apnea **Task**-no **Practitioner**-knowledge obstructive and central apnea and differential diagnosis limited **Team**-Communication and coordination activities did not exist **Working cond**-no **Organization**-Sedation rescue not standardized **Political**-no

management of rare events. Modifying an available pediatric human simulator allowed scenarios to be produced by the simulator when calibration protocols were followed. The simulator could be moved significant distances within our facility and could be interfaced with the actual devices used by clinicians in the settings in which it was evaluated. Physiologic data files and event logs were easy to synchronize to video, using the time that the sedation provider administered

Table 2. Qualitative review to identify rescue system problems in an actual sedation unit, cont.

Care Management Problems	Description	Contributory Factors
4. Advanced Management Errors A. Attending anesthesiologist did not respond because of STAT paging system failure B. Code team anesthesiologist did not cross check positive pressure ventilation C. Anesthesiologist attempted to "lead" but too engaged in task execution E. Differential never stated, naloxone never considered	STAT paging systems unknown to new anesthesia attending. Never realized they were needed STAT. The code team attending anesthesiologist did arrive and went to the head of the bed. They initially took over positive pressure ventilation. However, while receiving a report on the situation, they started requesting equipment for intubating and stopped ventilating. During this time a respiratory therapist arrived. Both the attending anesthesiologist and respiratory therapist did not optimize positive pressure ventilation. Both left the patient apneic while gathering materials to intubate. During this entire time the SpO_2 worsened.	**Patient**-Apnea **Task**-Code dispatching failure **Practitioner**-Attentional error and strategic error. Assumed RN had optimized positive pressure ventilation. Did not cross-check ability to provide positive pressure ventilation. Did not ventilate when getting read to intubate **Team**-No leader **Working cond**-noisy due to over 10 people in room, many talking all at once **Organization**-no **Political**-no
5. Crew Resource Management Errors A. Team structure (multiple responders dispatched that were not needed) defined by overloaded leader B. Team situational awareness poor C. Team goals, role clarity, communication, and coordination poor	The primary personnel of the RN, PCT, and code team did not function. An excess of personnel was present. However, their roles were ill defined, communication was poorly directed. When requests made, multiple individuals responded. Considerable delays in task execution (getting equipment for intubation took several minutes).	**Patient**-no **Task**-no **Practitioner**-Leadership and "followership" poor **Team**-Team structure defined dynamically, leadership poor, communication poor **Working cond**-crowding by individuals that did not contribute **Organization**-limited code team training **Political**-no

sedative medication as the scenario start time. A single wide-angle-view videotape proved adequate as a data source and allowed for post-hoc analysis by multiple experts. The audio provided by the digital video camera was also adequate for analysis, but the amount of noise made it difficult to follow individual commentary.

This methodological study investigates the use of provocative testing in the sedation care domain as a means for objectively measuring safety. Specifically, we were interested in the measurement of pediatric rescue capability. The method we describe has great potential for achieving this aim. Based on a single sedation care system test, we were able to show that the sedation system currently in use had multiple latent system failures. The rescue efforts deviated significantly from gold-standard care. The actual rescue was associated with more than 3 minutes of

hypoxia and 90 seconds of severe hypoxia. The first responders, despite having met training requirements for sedation, did not stimulate the patient when respiratory effort was initially still present. They also were delayed in selecting the correct size of oral airway and inserting it. The STAT paging system failed to mobilize the attending anesthesiologist covering this sedation location. The code team members responsible for advanced airway management assumed that the initial airway management was adequate and failed to assume care to get a positive outcome.

Our observations have been that the skill set associated with positive pressure ventilation is quite complex and that anesthesiologists and respiratory therapists are far more successful at creating a seal and ventilating than are novices, who rarely perform this task. In this case, the experts never tried to provide positive pressure ventilation (a strategic error). Many other latent system failures were identified (Table 2) and these findings proved to be robust and easy to corroborate. For example, subsequent tests of our paging system were associated with the same failure in six of eight tests. This latent system failure has since been corrected.

The qualitative analysis performed used the approach described by Vincent[15] to identify care management problems and associated contributing factors. These latent failures existed in an actual clinical unit that was providing sedation care. This provocative test uncovered failure opportunities across the full spectrum of resources supporting sedation care—from the blunt-end (system resources) to the sharp-end (provider-interface factors) of the system.[21]

To date, the METI human simulators have been used primarily as a tool for training. Training applications range from teaching basic skills to high-order crisis resource allocation and team collaboration.[22] This study supports the finding that, even in training exercises, latent system failures can be uncovered while managing a simulated event. DeAnda and Gaba[23] identified 132 unplanned incidents during 19 simulations using a modified critical incident methodology (range = 3–14 incidents, mean = 6.9 incidents per simulation exercise). In addition, the classes of incidents were similar to those identified previously by Cooper et al. in field studies.[24, 25]

In summary, this study demonstrates that using simulation as a safety probe is feasible and of enormous potential impact. In addition to its practical aims, this work seeks to exemplify the type of research that increasingly will be needed in efforts to improve patient safety. The field experiment methodology used here represents one such approach that human factors researchers have employed successfully in a number of domains. These researchers have pointed out[13, 26–28] that, to make research relevant to practical problems in complex work systems, naturalistic studies are a necessary complement to traditional, controlled experimental research. Future investigations using this methodology will benefit from a more sophisticated scenario and must be validated against empiric data. Additionally, video coding schemes that delineate ideal performance and allow for the easy identification of performance deviations need to be refined. While these investigations are being pursued to advance the efficacy and safety of

pediatric procedural sedation, our ultimate goal is to provide a fundamental method for assessing safety in domains in which events are rare, but doomed to happen. Only by designing our health care systems not only to use preventive strategies, but strategies that capture error and support recovery, will safety be maximized.

Conclusion

This research supports the feasibility of using available human simulation as a crash-test dummy, capable of measuring the rescue systems used in actual sedation-care settings. This provocative test readily allowed rescue performance in different sedation-care settings (such as the emergency room and in radiology departments) to be contrasted with the relevant gold-standard performance. We also demonstrated that personnel deemed competent and safe on the basis of meeting our hospital training requirements for airway management had profound performance deviations when compared to gold-standard practice.

Acknowledgments

Funded in part by NIH grant 1 RO3 HD041229-01, National Institute for Child Health and Human Development, and the Burnap Fund for the PainFree Pediatric Simulator.

Author affiliations

Dartmouth-Hitchcock Medical Center, Lebanon, NH (GB, JC, SA, JJ). Cognizant Systems, Toronto, Ont. (KC).

Address correspondence to: George Blike, M.D., associate professor of Anesthesiology, Dartmouth Medical School, One Medical Center Drive, Lebanon, NH 03756; phone: 603-650-2908; e-mail: Blike@hitchcock.org.

References

1. Kohn LT, Corrigan, JM, Donaldson MS, editors. To err is human: building a safer health system. A report of the Committee on Quality of Health Care in America, Institute of Medicine. Washington, DC: National Academy Press; 2000.
2. Blike G. Cravero J. Nelson E. Same patients, same critical events—different systems of care, different outcomes: description of a human factors approach aimed at improving the efficacy and safety of sedation/analgesia care. Qual Manag Health Care 2001;10(1):17–36.
3. Holloway VJ, Husain HM, Saetta JP, et al. Accident and emergency department led implementation of ketamine sedation in pediatric practice and parental response. J Accid Emerg Med 2000 Jan;17(1):25–8.
4. Petrack EM, Christopher NC, Kriwinsky J. Pain management in the emergency department: patterns of analgesic utilization. Pediatrics 1997;99:711–4.
5. Lowrie L. Weiss AH, Lacombe C. The pediatric sedation unit: a mechanism for pediatric sedation. Pediatrics 1998 Sept;102(3):E30.
6. Havel CJ Jr., Strait RT, Hennes H. A clinical trial of propofol vs midazolam for procedural sedation in a pediatric emergency department. Acad Emerg Med 1999 Oct; 6(10):989–97.
7. McCarty EC, Mencio GA, Walker LA, et al. Ketamine sedation for the reduction of children's fractures in the emergency department. J Bone Joint Surg AM 2000 Jul;82(A):912–8.

8. Egelhoff JC, Ball WS, Koch BL, et al. Safety and efficacy of sedation in children using a structured sedation program. Am J Roentgenol 1997 May;168 (5):1259–62.

9. Heuss LT, Schnieper P, Drewe J, et al. Risk stratification and safe administration of propofol by registered nurses supervised by the gastroenterologist: a prospective observational study of more than 2000 cases. Gastrointest Endosc 2003;57(6):664–71.

10. Epstein B. Department of Health and Human Services, Office of Epidemiology and Biostatistics, Center for Drug Evaluation and Research. Data Retrieval Unit HFD-737; June 27, 1989.

11. Cote C J, Notterman DA, Karl HW. Adverse sedation events in pediatrics: a critical incident analysis of contributing factors [comment]. Pediatrics 2000;105(4 Pt 1):805–14.

12. King PH, Pierce D, Higgins M, et al. A proposed method for the measurement of anesthetist care variability. J Clin Monitor Comp 2000;16(2):121–5.

13. Woods DD. Process-tracing methods for the study of cognition outside of the experimental psychology laboratory. In: Klein G, Orasanu J, Calderwood R, et al, editors. Decisionmaking in action: models and methods. Norwood, NJ: Ablex; 1993. pp. 228–54.

14. Vincent C, Taylor-Adams S, Chapman EJ, et al. How to investigate and analyze clinical incidents: clinical risk unit and association of litigation and risk management protocol. BMJ 2000;320:777–81.

15. Vincent C. Patient safety: understanding and responding to adverse events. NEJM 2003 Mar; 348(11):1051–6.

16. Blike GT, Unpublished data regarding quality assurance incident review in the Department of Anesthesia at Dartmouth-Hitchcock Medical Center; 2003 Oct 3.

17. Roth EM, Bennett K, Woods DD. Human interaction with an "intelligent" machine. Int J Man–Machine Studies 1987;27:479–525.

18. Cook RI, Woods DD, Miller C. A tale of two stories: contrasting views of patient safety. Chicago: National Patient Safety Foundation; 1998.

19. Mackenzie CF, Jeffries NJ, Hunter A, et al. Comparison of self reporting deficiencies in airway management with video analyses of actual performance. Hum Factors 1996;38(4):623–35.

20. Brunswik E. Perception and the representative design of psychological experiments. 2nd edition. Berkeley, CA: University of California Press; 1956.

21. Cook R, Woods D. Operating at the sharp end: the complexity of human error. Chapter 13, In: Bogner S, editor. Human error in medicine. Mahweh, NJ: Lawrence Erlbaum Associates; 1994.

22. Gaba D, Fish K, Howard S. Crisis management in anesthesiology. Philadelphia: Churchill Livingstone; 1994.

23. DeAnda A, Gaba D. Unplanned incidents during comprehensive anesthesia simulation. Anesthesia & Analgesia 1990 Jul;71(1):77–82.

24. Cooper JB, Newbower RS, Long CD, et al. Preventable anesthetic mishaps: a study of human factors. Anesthesiology 1978;49:399–406.

25. Cooper JB, Newbower RS, Kitz RJ. An analysis of major errors and equipment failures in anesthesia management: considerations for prevention and detection. Anesthesiology 1984;60:34–42.

26. Perrow C. Normal accidents: living with high-risk technologies. New York: Basic Books, 1984.

27. Reason J. Human error. New York: Cambridge University Press; 1990.

28. Hoffman RR, Woods DD. Studying cognitive systems in context: preface to the special issue. Hum Factors 2000;42(1):1.