Aus der Klinik und Poliklinik für Anästhesiologie

der Universität Würzburg

Direktor: Professor Dr. Dr. h.c. Norbert Roewer

Simulation To Establish Benchmark Outcome Measures

Inaugural - Dissertation

zur Erlangung der Doktorwürde der

Medizinischen Fakultät

der

Julius-Maximilians-Universität Würzburg

vorgelegt von

Matthias M. Kurrek

aus Uffenheim

Würzburg, 01 August 2015

Referent:	Prof. Dr. med. Peter Kranke Klinik und Poliklinik für Anästhesiologie Universitätsklinikum Würzburg Zentrum Operative Medizin Oberdürrbacher Straße 6 97080 Würzburg, Germany
Koreferent:	Prof. Dr. med Wolfram Voelker Stellvertreter des Klinikdirektors der Medizinischen Klinik und Poliklinik I Oberdürrbacher Straße 6 97080 Würzburg, Germany
Dekan:	Prof. Dr. Matthias Frosch

Tag der mündlichen Prüfung: 6 Februar 2017

Der Promovend ist Arzt

CONTENTS

SIMULATION TO ESTABLISH BENCHMARK OUTCOME MEASURES

INTRODUCTION

The history of aircraft simulators

The idea of simulation to prepare for difficult tasks goes back many centuries, when warriors attempted to prepare themselves for combat through the use of simple mannequins. However, our modern generation usually thinks about simulation in the context of aircraft simulation, and it is Edward Link who is credited with the development of modern day aircraft simulators.

Edward Link was the son of an organ manufacturer who was raised in Binghamton, in the state of New York. While his father would have liked for his son to show more interest in his musical business, young Edward always showed a passion for flying. In the 1940s flying was not an easy task, requiring substantial manual dexterity and practice, and was, of course, associated with significant expense. In order to make learning the so-called 'stick and rudder' skills more efficient, Edward constructed, using several of the pneumatic controls from the basement of his father's organ factory, a small model airplane that allowed at least rudimentary simulation of the basic flight controls – the birth of the first flight simulator.

Edward subsequently opened a flight school, in which he offered lessons in his flight simulator, but it was not until the early days of World War II when the potential benefits of this revolutionary concept were widely recognized. Stories tell that during a meeting among several military commanders at an airfield, severe fog made for almost impossible weather conditions and prevented many of the pilots from attending. Just as the meeting was about to be cancelled, the sounds of an approaching aircraft were heard and were followed by a spectacular landing of Edward in his aircraft. When asked by the surprised attendees how he had managed to land with such little effort, Edward took the opportunity to

describe his flight simulator, with which he had prepared himself to fly under various adverse conditions with limited visibility.

The concept of simulation was immediately embraced and led the military to order several of his simulators, and marked the beginning of one of the most successful flight simulation businesses in the history of aviation – CAE Link. Over the ensuing decades, aircraft simulation (while expensive) proved to be a viable and economically attractive supplement to learning by flying in a real aircraft, the cost of which exceeds simulation by a magnitude.

The beginning of cockpit resource management training in aviation

The next big advancement in simulation came in the 1970s. By that time a variety of aviation accident investigations had shown that it was primarily poor teamwork and poor management skills rather than simple technical errors that were at the root of many tragedies, such as the infamous accident of Eastern Airlines flight 401, which crashed in December 1972 in the Florida Everglades after the cockpit team became distracted by a burned out light bulb, killing 99 people on board. In the aftermath of these investigations, Delta Airlines introduced the concept of Crew Resource Management (CRM), based on the business model of practicing teamwork and communication under stress. During CRM sessions, the entire team of a cockpit would practice the management of various crises in an aircraft simulator, followed by a video-assisted debriefing. While the pilots were initially skeptical about this concept, it was ultimately widely embraced, certainly in part through the pilots' obvious interest in being best prepared for any critical incidents and staying alive – after all, pilots go down with their planes…

Early simulators and crisis resource management in medicine

Simulation in medicine is not that new either – the earliest interest in the use of a computer-controlled mannequin goes back to the University of California, where in the 1960's Abrahamson and Denson constructed what is now probably known as one of the first simulators in medicine.[1] Unfortunately the computer technology was not yet at a point where it could support many of the mannequin's functions and vital sign displays, and the idea was not pursued.

In the 1970's, more or less independently, the idea of using a mannequin as a simulator in medicine resurfaced from two academic anesthesia centers in the US, with a slightly different focus. The University of Gainesville in Florida (under the direction of Dr. Michael Good) developed a full size mannequin that was controlled by a computer with a sophisticated software program to simulate respiratory physiology as well as pharmacokinetics and -dynamics. This model, marketed initially by Loral™ (followed by METI™ and then by CAE Healthcare) focused on teaching basic anesthetic skills.[2]

The second, and slightly different approach, originated at Stanford University where Dr. Gaba, an anesthesiologist and holder of a pilot's license, always wondered why anesthesiologists, often compared with the pilots ('hours of boredom – moments of terror') did not have the benefit of simulated crisis resource management training. He began to develop the idea of an Anesthesia Crisis Resource Management Course (ACRM), by putting a (fairly simple) mannequin in a simulated OR with actors, while controlling the vital sign display manually from a separate computer.[3] His mannequin was soon after equipped with the computer models developed by Dr. Howard Schwid (Department of Anesthesia, University of Washington), and marketed by CAE Link (this line of simulators was later discontinued).

Within only a short time simulation centers opened in Canada, Europe and the rest of the world and departments of anesthesia took the lead to teach ACRM courses in simulation centers.[4,5] Besides being used to teach ACRM, simulation began to be used as a powerful research tool, including studies on

the use of simulation for evaluation,[6,7] assessment of the effect of sleep deprivation[8] as well as an educational tool.[9]

While simulation in aviation proved financially to be a very attractive alternative when compared to learning in real aircrafts, the same can unfortunately not be said for realistic simulation in medicine[10] and the majority of centers require ongoing support from educational institutions. Despite the expense, the use of simulation has increased dramatically and represents one of the most significant advances in medical education of the last decades.

Performance evaluation using realistic simulation

Miller has described performance as a pyramidal concept, involving elements of knowledge ('knows'), the application of knowledge into a care plan ('knows how'), the implementation of that plan ('shows how') and finally the actual clinical practice ('does').[11] Traditional evaluations are usually based on demonstration of theoretical knowledge (written/oral test), together with observation of 'routine' care, limiting the assessment of key areas related to the practical management of critical events (crisis management).

With the arrival of realistic simulation the interest to use simulated critical events to assess performance grew rapidly. It was now possible to reproduce clinically challenging scenarios - without putting a patient at risk - in order to evaluate the performance of various participants. Early studies focused mostly on technical markers (i.e. 'was a certain action performed or not?'), which could be measured fairly easily with good reliability and which showed different levels of performance among various groups of practitioners.[12,13,14] This led to efforts to define the role of simulation as part of the formative assessment[15] and to integrate simulation-based accreditation into the National Board Examination.[16]

Assessment of non-technical skills in simulation

It had been widely recognized that human factors (as opposed to simple technical errors) are frequently the root cause for human error in medicine and a number of tools, some of which were adapted from the aviation industry, were developed in order to measure behavioral and cognitive performance in a simulator.[17,18] However, the assessment of behavioral markers (such as, for example, the concept of 'situational awareness') is more difficult to quantify, associated with significant inter-rater reliability and may require a substantial number of reviewers/raters in order to achieve reproducible assessments, but this ability to quantitatively measure cognitive performance allows the impact analysis of educational interventions geared to improve human factors and ultimately human errors.[19,20,21,22]

Expanding to team training with realistic simulation

Following the landmark publication by the Institute of Medicine "To Err is Human",[23] estimating that over 50,000 patients per year may die from human error, the Institute of Medicine recommended that 'health care organizations should establish interdisciplinary team training programs for clinicians to incorporate the proven team training strategies used in the aviation industry'.[24] However, up to now simulation-based education had been applied and its impact been evaluated only to individual health care practitioners. It now became necessary to broaden the concept and to design and test an evaluation tool for an entire health care team as a first step to designing and evaluating the impact of an educational module.

In the United Kingdom communication had been identified as a considerable problem by the Confidential Enquiry into Maternal and Child Health (CEMACH)[25] and 'skill drills' became a requirement in the new Maternity Clinical Negligence Scheme for Trusts (CNST).[26] Using interdisciplinary obstetrical teams, Morgan et al developed a reliable and valid performance tool for the

Assessment of Obstetrical Team Performance (AOTP) and Global Assessment of Obstetrical Team Performance (GAOTP)[27] that can be used to evaluate the impact of an educational intervention.

The dilemma of defining clinically acceptable performance

Due to the rare occurrence of many critical events, it is usually difficult to gather real life data of sufficient quality and quantity to decide if the management of an actual event fell below what could reasonably be expected of a peer. However, the question whether an intervention has been carried out in a timely manner can be relevant for quality assurance, the design of educational curricula (and also as part of the defense against alleged substandard care).

One approach could be to collect data from realistic crisis simulations and try to define what constitutes an outlier among various performances by plotting the outcome variable in question (for example the time from asystole to chest compression) from a number of different teams. Analogous to the manufacturing industry, one could ultimately attempt to set up tolerance limits and define normality accordingly; however, this would require very large sample sizes.[28]

We intended to evaluate the use of simulation data to describe the distribution of management times as an approach to decide if the team management of a simulated obstetrical crisis scenario could be considered an outlier.[29]

METHODS

Institutional ethics review and approval was obtained by Sunnybrook Health Science Centre, University of Toronto Research Ethics Board (Chairperson: Dr. Philip Hebert, 2075 Bayview Ave, Toronto, Ontario M4N 3M5, Canada, approval REB# 351-2006, October 2006) and each participant provided written consent. The subjects of this study were previously reported as part of a project to determine the psychometric properties of a behavioral marking system for obstetrical team training in a high-fidelity simulator.[27] The data have been uploaded to datadryad.org (data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.8s511). The simulation facility consisted of a realistically-equipped hospital room with all relevant equipment available, including anesthesia gas machine, and used the SimMan™ (Laerdal Medical Canada, Ltd., 51 Nashdene Road #45, Toronto, ON M1V 4C3 Canada) full-scale realistic mannequin with various computer controlled features (voice, anatomically correct airway, heart and breath sounds, etc.). An add-on module was specifically designed and built for the obstetrical scenarios and consisted of a pregnant abdomen with a simulated amniotic sac through which the baby (or babies) had to be delivered via cesarean section. A fetal heart rate simulator provided information on both contractions and fetal heart rate tracings. A pump was installed that could simulate a massive obstetrical hemorrhage.

Four obstetrical simulation scenarios were developed using morbidity and mortality data from the UK Centre for Maternal and Child Enquiries (CMACE). The four scenarios (see Table 1), which had previously been developed included: A) need for Cesarean section under general anesthesia with a difficult airway, can't intubate/can't ventilate resulting in hypoxia and leading to pulseless electrical activity; B) severe pre-eclampsia, epidural in situ, non-reassuring fetal heart rate tracing leading to urgent cesarean section and development of pulmonary edema; C) 34-week twin gestation umbilical cord prolapse, emergency cesarean section complicated by amniotic fluid embolism and asystole; D) prolonged fetal bradycardia with emergency cesarean section, occult abruption and massive bleeding.

For the study, 12 multidisciplinary teams (one specialty-certified staff obstetrician, one specialty-certified staff anesthesiologist, 3 staff obstetrical nurses and in some cases, depending on the team's usual practice pattern, a family doctor) managed all 4 scenarios. The participants' age (mean±SD) and years in practice (mean±SD) were as follows: nurses (n=36): 40.4±10.7 and 15.8±11.5; family MDs (n=6): 37.4±6.4 and 8.6±4.7; obstetricians (n=12): 47.0±10.0 and 14.5±10.1; anesthesiologists (n=12): 41.0±7.3 and 7.4±5.5. Each of the teams was recruited from one of the various local teaching or community hospitals (where they usually worked with each other) and had not received previous high-fidelity simulated obstetrical team training. Every team received a thorough instruction and tour of the simulation centre. At the beginning of each scenario the person to first enter the patient room was given a detailed history of the patient, results of the physical examination with results of pertinent laboratory and the opportunity to interview the patient before scenario begin. Further details were available in the patient's chart by the bedside. The simulation operator was the same person for all scenarios and the principal investigator oversaw the scenario from the control room with the simulation operator to ensure consistency of presentation. The principal investigator (based on clinical judgment) determined and instructed the simulation operator when to stop the scenario.

Clinically relevant outcome variables for each scenario were defined by a multidisciplinary group of obstetricians, anesthetists and nurses with many years of clinical experience in obstetrics (see Table 1). All sessions were videotaped and time to performance of the outcome measure recorded in seconds by a trained observer who was unaware of the identity of any individual/team.

The times from each team who performed the outcome measure in questions were used to calculate the median and quartiles with their associated 95% confidence intervals. The confidence limits were distribution free and calculated using rank order statistics. All analyses were carried out using SAS Version 9.2 (SAS Institute, Cary, NC, USA).

Scenario		Outcome variable
Scenario A:	need for Cesarean section under general anesthesia with a difficult airway, can't intubate/can't ventilate resulting in hypoxia and leading to pulseless electrical activity	A1: time from induction to delivery
		A2: time from absence of saturation to initiation of CPR
Scenario B:	severe pre-eclampsia, epidural in-situ, non-reassuring fetal heart rate tracing leading to urgent cesarean section and development of pulmonary edema	B: time from rapidly decreasing saturation to fall below 90 to decision to intubate
Scenario C:	34 week twin gestation umbilical cord prolapse, emergency cesarean section complicated by amniotic fluid embolism and asystole	C1: time from onset asystole to initiation of CPR
		C2: time from onset of asystole to administration of epinephrine
		C3: time from verbalization for need to perform cesarean section to delivery of babies
Scenario D:	prolonged fetal bradycardia with emergency cesarean section, occult abruption and massive bleeding	D: time from decision for cesarean section until delivery

Table 1. Clinically relevant outcome variable for each scenario were defined by Delphi method

RESULTS

All twelve teams completed all four scenarios (see Figures 1-4.) We were able to calculate the median, 75th and 90th percentiles for time to completion for all 7 clinical outcomes (Table 2) using non-parametric methods to account for non-normally distributed data. The confidence intervals, given the small sample size, were large.

Task C1 (time from onset asystole to initiation of CPR during amniotic fluid embolism) was found to be the shortest one to completion (median 66s [95% CI 26-97s]; 75th quantile 89s [95% CI 86-97s]), whereas task D (time from decision for cesarean section until delivery during massive bleeding) took the longest to complete (median 359s [95% CI 350-426s]; 75th quantile 398s [95% CI 361-426s]).

One outcome variable (Scenario A - cannot intubate, cannot ventilate arrest – outcome A2: time from absence of saturation to initiation of CPR) was only done by 7 of the 12 teams before scenario end.

Scenario Outcome variable (number of teams that performed action)	50th quantile (lower-upper 95% CI in s)	75th quantile (lower-upper 95% CI in s)	90th quantile (lower-upper 95% CI in s)
A1 (11/12)	312s (249-389)	354s (276-917)	389s (354-917)
A2 (7/12)	215s (89-384)	250s (194-384)	384 (221-384)
B (11/12)	79s (57-215)	134s (65-407)	215s (134-407)
C1 (12/12)	28s (14-86)	66s (26-97)	89s (86-97)
C2 (11/12)	93s (63-167)	124s (89-197)	167s (124-197)
C3 (11/12)	304s (275-350)	349s (287-393)	350s (349-393)
D (12/12)	350s (316-361)	359s (350-426)	398s (361-426)

Table 2. Outcome measure for the twelve teams in the four scenarios

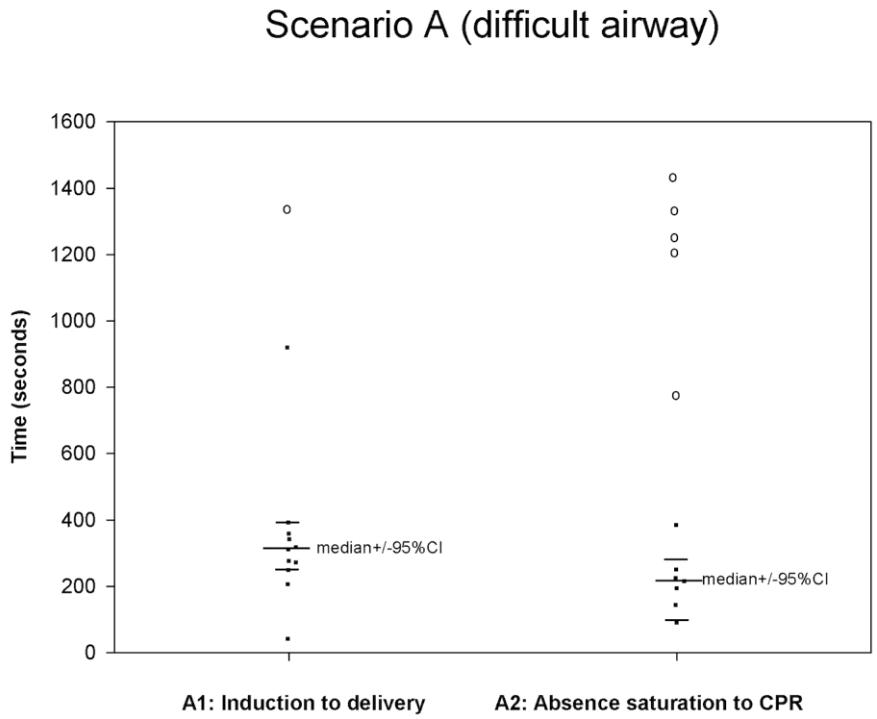

Figure 1.

Figure 1. Scenario A. The median was calculated using teams performing the outcome measure (teams not performing the outcome measure are represented by open circles placed at the time when their scenario was stopped).

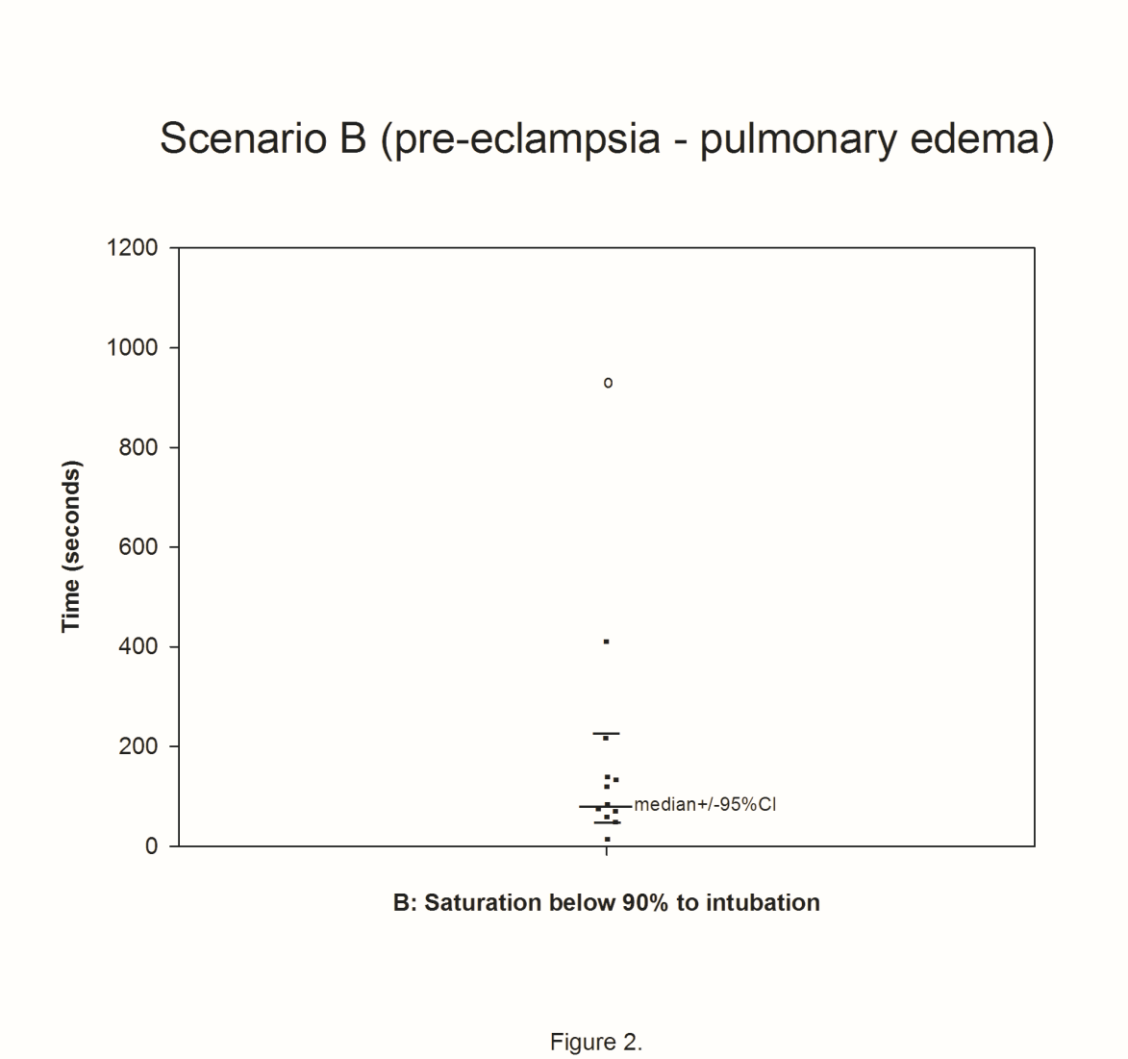

Figure 2.

Figure 2. Scenario B. The median was calculated using teams performing the outcome measure (teams not performing the outcome measure are represented by open circles placed at the time when their scenario was stopped).

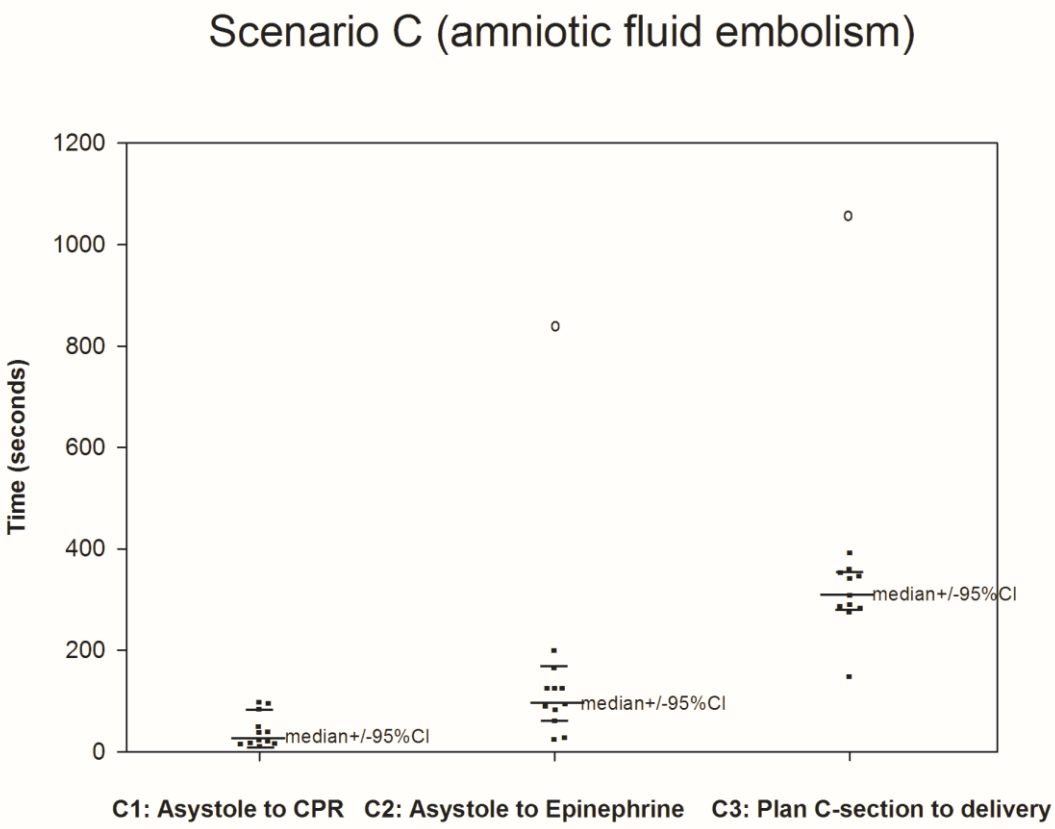

Figure 3.

Figure 3. Scenario C. The median was calculated using teams performing the outcome measure (teams not performing the outcome measure are represented by open circles placed at the time when their scenario was stopped).

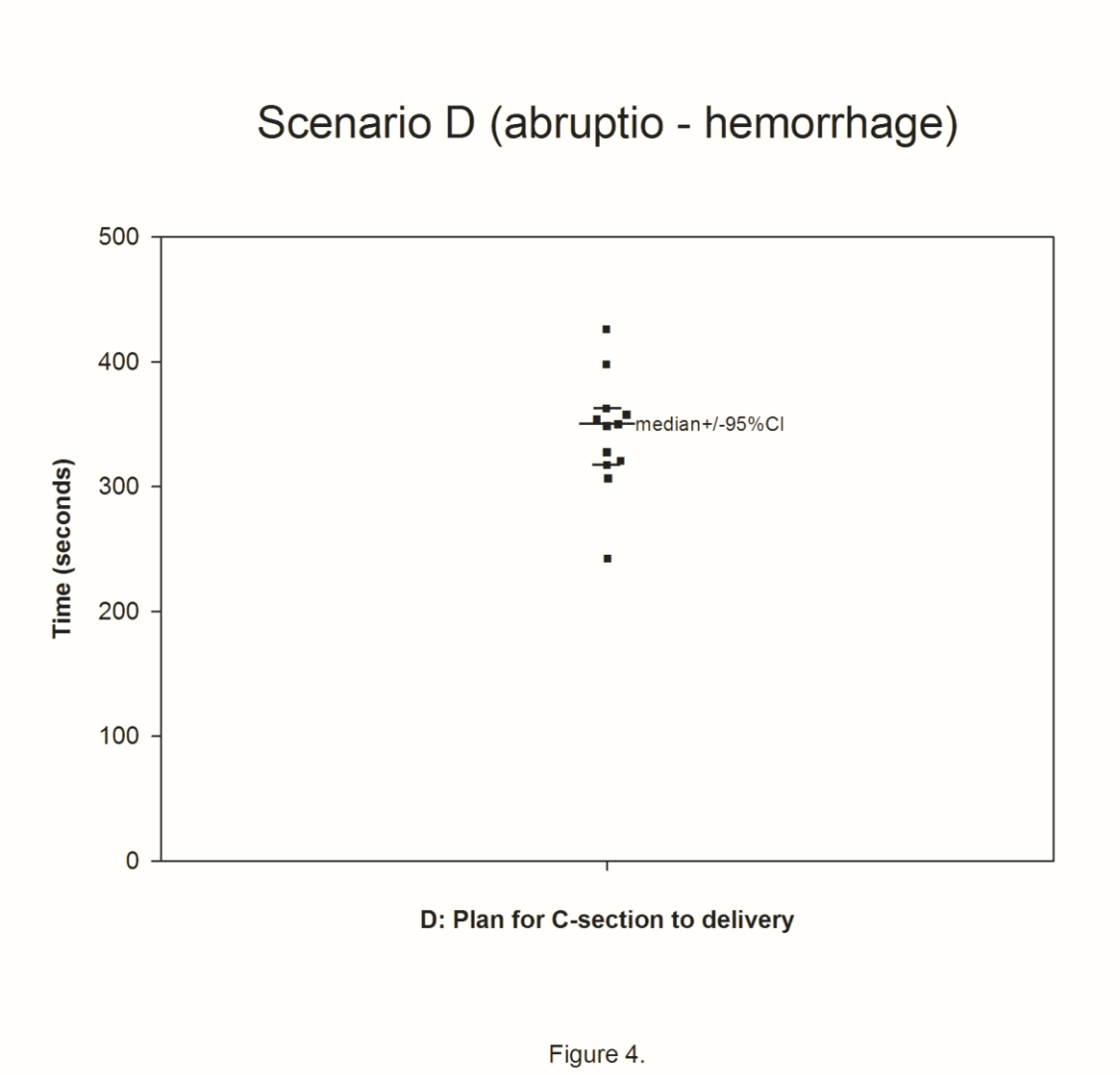

Figure 4.

Figure 4. Scenario D. The median was calculated using teams performing the outcome measure (teams not performing the outcome measure are represented by open circles placed at the time when their scenario was stopped).

DISCUSSION

Time is often one of the few measures that can be extracted from clinical records with at least a moderate degree of certainty following a critical event. There are very few expert opinions on a national level which define expectations (such as the recommendation to perform a perimortem cesarean delivery within 4 minutes of maternal cardiac arrest, as introduced in 1986)[30] and clinicians often have very little objective data to support or refute a claim that an action was so late that it fell below what could reasonably be expected of a peer. We have demonstrated the use rank order statistics to calculate quantiles with confidence limits for management times of critical obstetrical events using data from realistic simulation. This approach could be used to yield information that may assist in the decision if a given performance could be considered normal and could also point to aspects of care that seem to pose particular challenges as evidenced by a large number of teams not performing the expected maneuver.

Ideally, large clinical databases could provide information about the management of critical events, but this is usually not feasible, particularly for events that occur with very low frequency. Realistic simulation, on the other hand, has become very common and allows the exposure of a large number of clinical practitioners to similar critical events in a high fidelity environment.[31] The authors know of only one clinical registry - the recently established 'Get-with-The-Guidelines Resuscitation' registry from the American Heart Association - that captures real life cardiac arrest management times (only). The results of such a registry would allow an interesting validation of data from simulated arrests.

Early studies have focused on the use of realistic simulation to evaluate individual technical and non-technical skills[6] whereas more recent works have begun to explore work with entire medical teams,[32] recognizing the central role that team performance plays for patient safety. In the context of individual as well as team performance, especially for summative evaluations,[33] the question of threshold performance naturally arises: which result is deemed to be acceptable

and which result is deemed to be substandard. However, establishing credible and accepted cut-off points for performance during summative examinations can be challenging, often relying on subjective expert opinion.[34] The assessment of performance for entire medical teams adds significant challenges that remain to be resolved, such as issues with reliability and relative contributions of individuals versus team skills for overall performance.[35] Most results of individual simulated performances have been validated by comparing different groups of practitioners with various training backgrounds (medical students versus residents versus staff etc.), using scoring templates that are often based on consensus about which action is deemed appropriate (i.e. action 'x' must be performed in order to score a point).[36] The information generated from these scores is therefore limited when trying to evaluate the management of certain critical incidents in real life.

Rather than defining a standard in absolute terms, we attempted to use rank order statistics to calculate quantiles with confidence limits for management times from a small data set of realistic simulations as a possible approach to decide the difference between acceptable and not acceptable, analogous to the concept of defining normality using the Gaussian distribution. This approach of defining normal in a statistical sense is indeed very common for the purpose of creating reference values through tolerance limits especially in manufacturing and laboratory medicine.[28, 37]

Since the current calculations were done for only 12 teams, the resulting confidence intervals around the calculated median response times and the quantiles were very large. However, with the pooling of results from several simulation centres the amounts of data for these calculations could be significantly expanded and this would provide more acceptable confidence intervals for those parameters as well as make the calculation of tolerance limits possible. A large database of outcomes may also allow the establishment of a 'grey zone' of time-based intervention values from the reference group and thus treat acceptability as a continuous, rather than as a binary (acceptable vs. unacceptable) variable.

It may be of interest to create a central data bank of simulation results that could ultimately be used to make those results available to interested users

(analogous to an anesthesia registry) and to create reference data for the skills of both teams as well as individual practitioners from various backgrounds during different scenarios. However, if results from several simulations are to be pooled, then it will be necessary to ensure that the scripting of the scenarios is comparable.

It has to be remembered that the tolerance limits from such data would be derived from realistic simulations and this should of course be kept in mind when trying to make inferences on clinical performance. While the realism of case presentations and simulated incidents is generally rated as fairly high among participants[3], most participants anticipate that something will happen during the simulation and may thus display heightened vigilance and faster response times when compared to real life. Also, especially with respect to invasive procedures (including cesarean sections) the threshold to perform such an intervention and the time to complete the procedure may not necessarily be the same as in real life.

While time is often an easily accessible measure, it may not always be a good indicator of competent or even optimal performance. Time metrics may work well when individuals or teams are presented with situations in which single diagnoses and/or treatments are equally plausible at the beginning. When there are multiple possibilities and the "real one" isn't known until more information is collected, practitioners who - just by “lucky guessing” - will have a short response time even though their decision-making or problem solving may not be ideal. On the other hand, someone who choses something equally valid at first would be assigned a longer response time (and ‘worse’ performance) even though they both made an equally good first choice and the longer response time may indeed reflect a high degree of diligence. The most appropriate metric for those latter scenarios, taking in account decision-making and problem solving as markers of quality performance, may be a non-technical marker (perhaps in addition to time) and will have to be decided for each scenario in question.

Furthermore, there are likely variations in performance that will lead a clinician to a good outcome and vice versa. The “standard of care” defines what a reasonably prudent medical provider would or would not have done under the

same of similar circumstances. Currently, little is known about how a reasonable and prudent practitioner performs his or her work. Simulation might allow a window (though not perfect) into this very important area.

An outcome measure that is not performed by a significantly large number of participants before the end of the scenario may reveal particularly challenging clinical encounters for practitioners. While the commencement of CPR after the loss of the saturation signal (indicating a pulseless electrical activity arrest following a cannot intubate/cannot ventilate condition) was considered an essential action and tracked as part of the study, it was not done by 41% of the participating teams. This was surprising and while such a finding could result from a lack of realism of the scenario, the participants in our study did not indicate that they found the scenario unrealistic. If this is consistently observed in similar simulations, then this type of observation is extremely important because it would allow planning educational interventions in order to address what may be perceived as shortcomings in clinical management.

SUMMARY AND CONCLUSION

Following the early experiences in aviation, medical simulation has rapidly evolved into one of the most novel educational tools of the last three decades. In addition to its use in training individuals or teams in crisis resource management, simulation has been studied as a tool to evaluate technical and non-technical skills of individuals as well as, more recently, entire medical teams.

It is usually fairly difficult to obtain clinical reference data from critical events to refute claims that the management of actual events fell below what could reasonably be expected and we demonstrated the use of rank order statistics to calculate quantiles with confidence limits for management times of critical obstetrical events using data from realistic simulation. This approach could be used to describe the distribution of treatment times in order to assist in deciding what performance may constitute an outlier. It can also identify particular challenges of clinical practice and allow the development of educational curricula. While the information derived from simulation has to be interpreted with a high degree of caution for a clinical context, it may represent a further 'added value' or important step in establishing simulation as a training tool and to provide information that could be used in an appropriate clinical context for adverse events. Large amounts of data (such as from a simulation registry) would allow the calculation of acceptable confidence intervals for the required outcome parameters as well as actual tolerance limits.

ZUSAMMENFASSUNG

Im Anschluss an die Entwicklung der Simulation in der Flugtechnik hat sich die Simulation in der Medizin rasch innerhalb der letzten dreißig Jahre als eine der revolutionierendesten Methoden der Forschung und Lehre entwickelt. Ihre Benutzung hat sich während dieses Zeitraumes erheblich weiterentwickelt: anfänglich in erster Linie für Lehre des persönlichen (und Team) Management von medizinischen Notfällen und inzwischen auch für die standardisierte und

objektive Bewertung von persönlicher (und Team) Kompetenz von technischen und auch ‚nicht-technischer' Komponenten.

Es ist auf Grund der Rarität von vielen Notfällen normalerweise nicht möglich genug klinische Daten zur Auswertung zur Verfügung zu haben, um sagen zu können, ob das Management eines bestimmten Falles innerhalb von ‚normalen' Grenzwerten fällt. In dieser wissenschaftlichen Arbeit zeigten wir das ‚Rank Order Statistiks' dafür benutzt werden könnten, die Resultate von simulierten Notfällen in der Geburtshilfe als Bandbreite von ‚normalen' klinischen Leistungen darzustellen. Dieses Vorgehen würde es erlauben, eine klinische Leistung mit einer Datenbank von vergleichbaren simulierten Zwischenfällen abzugleichen, um entscheiden zu können, ob die klinische Leistung innerhalb von ‚normalen' Werten ausgefallen ist. Dieses Vorgehen verschafft außerdem Einblick, welche Probleme besondere Schwierigkeiten bereiten sodass ggf. gezielte Fortbildungen vorbereitet werden könnten. Obwohl die Daten der Simulation mit gewisser Vorsicht zu interpretieren sind, repräsentiert dieses Vorgehen eine neue Anwendung von Simulation, die für die Auswertung von klinischen Notfällen von großer Bedeutung sein könnte. Es wird in diesem Zusammenhang allerdings notwendig sein, relativ große Datenbanken von vielen simulierten Notfällen zu erstellen und auszuwerten, um die gesuchten Werte mit genug Genauigkeit kalkulieren zu können.

REFERENCES

[1] Abrahamson S, Denson J, Wolf R. Effectiveness of a simulator in training anesthesiology residents. *Journal of Medical Education* 1969; 44(6):515-519

[2] Good ML, Gravenstein JS. Anesthesia simulators and Training Devices. *International Anesthesiology Clinics* 1989; 27(3):161-168

[3] Gaba DM, DeAnda A. A Comprehensive Anesthesia Simulation Environment: Re-creating the Operating Room for Research and Training. *Anesthesiology* 1988; 69:387-394

[4] Kurrek MM, Fish KJ. Anaesthesia crisis resource management training: an intimidating concept, a rewarding experience. *Canadian Journal of Anaesthesia* 1996; 43(5Pt1):430-434

[5] Chopra V, Gesink B, DeJong J, Bovill J, Spierdijk J, Brand R. Does training on an anaesthesia simulator lead to improvement in performance? *British Journal of Anaesthesia* 1994; 73(3):293-297

[6] Gaba D, Howard S, Flanegan B, Smith B, Fish K, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral markers. *Anesthesiology* 1998; 89(1): 8-18

[7] Devitt JH, Kurrek MM, Cohen MM, Cleave-Hogg D. The validity of performance assessments using simulation. *Anesthesiology* 2001;95(1):36-42

[8] Howard SK, Gaba DM, Smith BE, Weinger MB, Herndon C, Keshavacharya S, Rosekind MR. Simulation study of rested versus sleep-deprived anesthesiologists. *Anesthesiology* 2003; 98(6):1345-1355

[9] Morgan PJ, Cleave-Hogg D, McIlroy J, Devitt JH. Simulation technology: a comparison of experiential and visual learning for undergraduate medical students. Anesthesiology 2002; 98(1):10-16

[10] Kurrek MM, Devitt JH. The cost for construction and operation of a simulation centre. *Canadian Journal of Anaesthesia* 1997; 44(11):1191-1195

[11] Miller G. The assessment of clinical skills/competence/performance. *Academic Medicine* 1990;65:S63

[12] Devitt J, Kurrek MM, Cohen MM, Fish K, Fish P, Noel AG, Szalai JP. Testing Internal Consistency and Construct Validity During Evaluation of Performance in a Patient Simulator. *Anesthesia Analgesia* 1998;86:1160-1164

[13] Forrest FC, Taylor MA, Postlethwaite K, Aspinall R. Use of a high-fidelity simulator to develop testing of the technical performance of novice anaesthetists. *British Journal of Anaesthesia* 2002;88(3):338-344

[14] Henrichs BM, Avidan MS, Murray DJ, Boulet JR, Kras J, Krause B, Snider R, Evers A. Performance of Certified Registered Nurse Anesthetists and Anesthesiologists in a Simulation-Based Skills Assessment. *Anesthesia Analgesia* 2009;108:255-262

[15] Salvodelli GL, Naik VN, Joo HS, Houston PL, Graham, M, Yee B, Hamstra SJ. Evaluation of Patient Simulator Performance as an Adjunct to the Oral Examination for Senior Anesthesia Residents. *Anesthesiology* 2006;104:475-481

[16] Berkenstadt H, Ziv A, Gafni N, Sidi A. The Validation Process of Incorporating Simulation-Based Accreditation into the Anesthesiology Israeli National Board Exams. *Israel Medical Association Journal* 2006;8:728-733

[17] Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Rating non-technical skills: developing a behavioural marker system for use in anesthesia. *Cognition, Technology & Work* 2004;6:165-171

[18] Kim J, Neilipovitz D, Cardinal P, Chiu M, Clinch J. A pilot study using high-fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients. The University of Ottawa Critical Care Medicine, High-Fidelity Simulation and Crisis Resource Management Study. *Critical Care Medicine* 2006;34:2167-2174

[19] Yee B, Naik VN, Joo HS, Salvodelli GL, Chung DY, Houston PL, Karatzoglou BJ, Hamstra SJ. Nontechnical Skills in Anesthesia Crisis Management with Repeated Exposure to Simulation-based Education. *Anesthesiology* 2005;103:241-248

[20] Welke TM, LeBlanc VR, Salvodelli GL, Joo HS, Chandra DB, Crabtree NA, Naik VN. Personalized Oral Debriefing Versus Standardized Multimedia Instruction After Patient Crisis Simulation. *Anesthesia Analgesia* 2009;109:183-189

[21] Morgan PJ, Kurrek MM, Bertram S, LeBlanc V, Przybyszewski T. Nontechnical skills assessment after simulation-based continuing medical education. *Simulation in Health Care* 2011;6(5);255-259

[22] Bruppacher HR, Alam SK, LeBlanc VR, Latter D, Naik VN, Salvodelli GL, Mazer CD, Kurrek MM, Joo HS. Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *Anesthesiology* 2010;112(4):985-992

[23] Kohn LT, Corrigan JM, Donaldson MS (eds). To Err Is Human – Building a Safer Health Care System. The Institute of Medicine, 2000 The National Academic Press, Washington DC

[24] Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC, Committee on Quality of Health Care in America, Institute of Medicine, 2001

[25] Sentinel event alert: Preventing infant death and injury during delivery. The Joint Commission, 1515 West 22nd Street, Suite 1300W, Oak Brook, Illinois 60523 USA 2004 http://www.jointcommission.org/sentinel_event_alert_issue_30_preventing_infant_death_and_injury_during_delivery/ (last accessed 30 June 2015)

[26] Brush L, Arulkumaran S. Clinical Negligence Scheme for Trusts for maternity services (CNST). *Current Obstetrics & Gynecology* 2003; 13:373-376

[27] Morgan PJ, Treguno D, Pittini R, Tarshis J, Regehr G, Desousa S, Kurrek M and Milne K. Determination of the psychometric properties of a behavioural marking system for obstetrical team training using high-fidelity simulation. *British Medical Journal Quality & Safety* 2012; 21:78-82

[28] Sasse EA. Determination of reference intervals in the clinical laboratory using the proposed guideline National Committee for Clinical Laboratory Standards C28-P. *Arch Pathol Lab Med* 1992; 116(7): 710-713

[29] Kurrek MM, Morgan P, Howard S, Kranke P, Calhoun A, Hui J and Kiss A. Simulation as a tool to collect outcome measures and to establish benchmarks. *PLoS One.* 2015 Jun 24;10(6):e0131064

[30] Katz VL, Dotters DJ, Droegemueller W. Perimortem cesarean delivery. *Obstetrics Gynecology* 1986; 68:571-576

[31] Gaba DM. Improving anesthesiologists' performance by simulating reality. *Anesthesiology* 1992; 76(4): 491-4

[32] Rosen MA, Weaver SJ, Lazzara EH et al. Tools for evaluating team performance in simulation-based training. *J Emerg Trauma Shock* 2010; 3(4): 353-9

[33] Berkenstadt H, Ziv A, Gafni N, Sidi A. Incorporating simulation-based objective structured clinical examination into the Israeli National Board Examination in Anesthesiology. *Anesth Analg* 2006; 102(3): 853-8

[34] Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teach Learn Med* 2006; 18(1): 50-7

[35] Manser T. Team performance assessment in healthcare: facing the challenge. *Simul Healthc* 2008; 3(1): 1-3

[36] Murray DJ, Boulet JR, Avidan M et al. Performance of residents and anesthesiologists in a simulation-based skill assessment. *Anesthesiology* 2007; 107: 705-13

[37] Sunderman FW. Current Concepts of 'Normal Values', 'Reference Values' and Discrimination Values' in Clinical Chemistry. *Clin Chem* 1975; 21(13): 1873-7

ACKNOWLEDGEMENTS

I appreciate the support and advice by **Prof. Dr. med Kranke** (“Doktorvater”) and **Prof. Dr. Voelker** (“Koreferent”) along with the assistance from the following individuals:

Ms. Mewis (Promotionsbüro Würzburg)

Prof. Dr. med. Roewer (Direktor der Anesthesiologie Würzburg)

Prof. Morgan (University of Toronto)

Prof. Howard (Stanford University)

Prof. Calhoun (University of Louisville)

Prof. Hui (University of California Los Angeles)

Dr. Kiss (University of Toronto)

Prof. Le Manach (McMaster University)

This work would not have been possible without the patience and encouragement of me dear wife, **Rubina,** and our children **Janine, Sebastian** and **Oliver**