Failure of simulation training to change residents’ management of oesophageal intubation


Department of Anesthesiology, Wake Forest University School of Medicine, Winston-Salem, NC 27157-1009, USA
*Corresponding author. E-mail: molympio@wfubmc.edu

Background. There are few scientific reports documenting the effects of simulation training on learning. Issues of scientific validity challenge investigators who measure such outcomes. We perceived a failure of residents to change their technical management of oesophageal intubation after simulation training and sought clarification of this observation.

Methods. Twenty-one residents were randomly exposed to two deliberate oesophageal intubation scenarios, first as a junior assistant (JS group) or as a senior managing resident (SS group), and secondly as a senior managing resident. After the first episode, residents were given an explanation and demonstration of the suggested technical management strategy, including: (i) confirmation of oesophageal intubation with a second direct laryngoscopy; and (ii) concurrent insertion of a second ETT (tracheal) tube. After the second episode, we retrospectively sought to confirm improvement in technical management within the SS group by measuring videotaped performances. Questionnaires were sent to the residents before and after reporting their performance results.

Results. There were 14 SS and seven JS subjects. Within SS, there was no improvement in ‘confirmation of oesophageal intubation with direct laryngoscopy’ (8/14 vs 9/14) or any improvement in ‘concurrent insertion of a second ETT (tracheal) tube’ (1/14 vs 2/14). Questionnaire responses offered considerable insight into these negative results.

Conclusions. This failure to change may have been secondary to a lack of criterion validity, lack of repetition or a long duration between episodes. The expectations for management were not regarded as being advantageous in simulation, but they were successfully adopted in actual clinical emergencies.

Br J Anaesth 2003; 91: 312–18

Keywords: education, simulation training; gastrointestinal tract, oesophageal intubation

Accepted for publication: April 10, 2003

Evolving over the past 20 yr, high-fidelity, full-scale patient simulation has emerged as an incredible learning opportunity that cannot be reproduced clinically. Whether or not learning actually occurs, however, is more difficult to demonstrate. A Medline and cross-reference search has located only two other scientific outcome studies1 2 and one testimonial report3 that detail the effects of high-fidelity simulation training on subsequent performance. Even the most recent study, by Byrne and colleagues,2 does not test the effects of crisis management training on the ability to subsequently respond to that same crisis. Furthermore, that study failed to demonstrate significant improvement in performance after videotape feedback. The study by Chopra and colleagues1 is the only such study to demonstrate a positive effect on the subsequent management of the same critical incident. One anecdotal report details several cases of the potentially life-saving value of human simulation, with a call for more anecdotal evidence.3

This article is accompanied by the Editorial.
Our group had developed concerns over the appearance of a lack of improvement in technical management of critical incidents, despite what was otherwise considered to be an effective training session. If an extensive amount of resources is expended on high-fidelity simulation, one would expect the best outcomes possible. Several other simulation studies document a surprisingly poor performance of experienced clinicians. DeAnda and Gaba,4±6 in a series of scholarly presentations, methodically characterize the human limitations of clinicians in recognizing and managing both latent and unplanned incidents. Their data show that ‘human error is ubiquitous, and that formal training and education should include recognition of events and the responses to them’5.

In one part of our residents’ simulation curriculum, we focus on the management of oesophageal intubation, a relatively common occurrence. Oesophageal intubation occurs even under controlled situations, and in patients with otherwise normal airway anatomy. At least two experts7,8 in crisis management have published suggestions about managing this event through (i) confirmation of oesophageal intubation by a second direct laryngoscopy, and (ii) placement of a second tracheal tube alongside the first tube, into the trachea while the first tube remains in the oesophagus. Especially when the oxygen saturation is maintained and immediate ventilation is unnecessary, this technique might otherwise prevent the aspiration of gastric contents and might facilitate visualization of the correct lumen. Furthermore, it proves beyond doubt that the original tube was indeed placed into the oesophagus.

Noticing a rather disappointing performance in managing subsequent oesophageal intubation, and considering the paucity of outcome studies in simulation, we were compelled to retrospectively analyse our data concerning this potentially devastating critical incident. This paper seeks to evaluate whether simulation training enhances the application of new technical skills, and why the learning process may have failed.

**Methods**

After approval of the Institutional Review Board, 25 CA-3, CA-2 and CA-1 (years 3, 2 and 1 respectively) anaesthesiology residents were asked to consent to a retrospective analysis of their performance in two separate human patient simulations, each of which contained a deliberate oesophageal intubation by a ‘junior’ (J) resident. The unknowing subject, a ‘senior’ resident (S), was expected to detect and manage the oesophageal intubation of this ASA class I patient. In each case, J described the laryngoscopic view as a grade I, with vocal cords easily visualized. The J subject pretended that he or she confidently placed the tracheal tube into the trachea while deliberately placing it into the oesophagus. All simulated patients had a Mallampati class I airway, and all presented with a need for rapid sequence induction. Attending supervision in each case was ‘not immediately available’.

Each of the two oesophageal intubation scenarios was only one part of a larger clinical scenario. One clinical scenario involved an otherwise healthy patient scheduled for appendectomy who developed anaphylaxis after resolution of the airway, while the other was a young and healthy trauma patient who developed haemorrhagic shock. There were other actors (a surgeon and/or a scrub nurse) randomly on site in approximately 25% of the scenarios. These actors had no defined role during the airway management of the patient. Each scenario was preceded with a verbal and/or written briefing of the case, and the scenarios were randomly assigned.

The random assignment of residents created two groups of test subjects, those who were senior in both sessions (SS), and those who were junior in the first session and senior in the subsequent session (JS). No resident was senior and then junior. There was a deliberate pattern of selecting CA-3 and CA-2 residents initially as seniors, paired with CA-2 and CA-1 residents as juniors. Junior subjects were then eligible to become senior subjects in the second session (Fig. 1).
Immediately after the first scenario, one of two investigators (MAO or RPAF) debriefed the S and J subjects together according to a prede®ned set of technical performance criteria derived from Schwid and colleagues8 (Table 1). We speci®cally emphasized two of these criteria as a novel, alternative approach to the oesophageal intubation: (i) a diagnostic second direct laryngoscopy to con®rm the oesophageal intubation, and (ii) concurrent placement of a second tracheal tube into the trachea. The logical sequence of approach was explained as follows: if (1) the glottic view was determined to be of grade I, (2) the oxygen saturation was maintained at 100%, (3) the patient was at risk of aspiration after insuf®lation of the stomach, and (4) a second direct laryngoscopy was required to intubate the trachea itself, then the resident was to (5) shift the oesophageal tube to the left side of the mouth, (6) perform the second direct laryngoscopy immediately with a second tracheal tube in hand, (7) insert the second tracheal tube into the trachea, and (8) use the first oesophageal tube as a conduit to suction the stomach with a nasogastric tube. Some, but not all, residents were allowed to practise this technique after the debriefing, time permitting.

Many of the residents (based on availability) who were involved in this scenario were then scheduled to return to the simulation laboratory as the senior resident in a different clinical scenario. The clandestine oesophageal intubation was repeated, initially out of curiosity of the principal investigator (MAO) and subsequently with an eye to retrospectively evaluating the adoption of the previously recommended procedure. After the second scenario, another debriefing was conducted emphasizing the same technical performance recommendations as before, but without specific reference to, or recall of, the residents’ performance in the first scenario. In fact, since this study was retrospective, at the time of the second performance we did not immediately have any written tabulation of their previous performance. It was the principal investigator’s perception (and disappointment) that a signi®cant lack of improvement in residents’ technical management had occurred, which prompted this retrospective evaluation.

Permission to proceed was obtained from the Institutional Review Board after ®ling a research study protocol. Residents who had completed either the JS or SS series were sent a request for informed consent to have their oesophageal intubation performances analysed. Enrolled subjects were then given a blinded pre-analysis question-
naire (Table 2) to determine their general thoughts about simulation training and their specific experiences and thoughts about managing oesophageal intubations. Two of the investigators (RPAF, ICMS) independently reviewed and scored the videotapes according to the equally weighted criteria in Table 1. The number of seconds between removal of the laryngoscope blade (after oesophageal intubation) and the first breath after tracheal intubation was documented. The principal investigator did not score the videotapes. The subjects’ prior simulation experience was
determined and the number of elapsed days between their first and second scenarios was counted. Blinded questionnaires were collected and tabulated, and data were analysed.

The overall results of the study and the subjects’ individual results were then reported to the subjects, and a post-analysis questionnaire (Table 3) was distributed; this was then returned and tabulated.

A Medical Education Technologies (Sarasota, FL, USA) human patient simulator, version 5.553, was used for this study. The mannequin was programmed to run as an equivalent to ‘standard man-awake’. Drug administration was scripted in the first scenario, but actually administered by the resident in the second. Choice of induction and/or paralytic agents was not allowed to affect the course of events. There were no adjustments made to the airway or pulmonary system; the pulmonary shunt was set to a minimum value to prevent arterial oxygen desaturation during the oesophageal intubation. Using a Videonics® Digital Video Mixer (FOCUS Enhancements, Inc., Campbell, CA, USA), two camera angles were intermittently selected for video recording, along with a third view of the haemodynamic monitor. Standard VCR with LCD image projection onto a SmartBoard® (SMART Technologies, Inc., Alberta, Canada) facilitated the videotape debriefings.

Statistical analysis
Tests for concordance between the two reviewers were conducted for each question and for the overall scores, and reported as a fraction of unity. Discordance for criteria 7 and 8 was verified by reviewing the original videotaped performances. All technical performance criteria were weighted equally.

Evaluation scores within each resident and test trial were averaged. The scores for questions 7 and 8 and total scores from the first simulation were compared with the scores of the second simulation using Fisher’s sign test.

Exact \( \chi^2 \) tests were then used to compare the two groups (SS vs JS) with respect to the second simulation outcomes for questions 7 and 8. Total second simulation test scores were compared between the two groups using the exact Wilcoxon rank sum test.

Spearman rank correlations were performed both overall and by group to test for any associations between prior experience and criterion 7 and 8 scores, and between total scores. The same correlations were performed to test for any associations between either interim days or seconds to ventilate and the change in scores between the first and second tests with respect to questions 7 and 8, and total scores. Finally, the same correlations were performed to test for any associations between either interim days or prior visits and the number of seconds to ventilate.

Results

Enrolment
Twenty-five residents were enrolled in the study and four were eliminated. Subject 2 (SS) was eliminated because the airway was mistakenly made difficult, and transtracheal jet ventilation was chosen above reintubation. Subject 24 (JS) inherently believed the airway to be difficult and chose a light-wand intubation. Subject 6 (SS) never had the initial oesophageal intubation. Subject 9 (SS) had a missing videotape of the first scenario. Of the 21 subjects remaining, there were 14 SS (four CA-3, five CA-2 and five CA-1 residents) and seven JS (two CA-2s and five CA-1s).

Concordance
Concordance was 0.93 overall, 1.00 for question 7, and 0.97 for question 8, which was secondary to a single discrepancy. The original reviewer corrected this error after reanalysing the video in question (which showed that the subject did not place a second tracheal tube), and the concordance for question 8 was changed to 1.00.

Total scores
Residents in group SS experienced a statistically significant improvement in overall total scores \( (P=0.0063) \) from the first to the second episode, ranging from a decrease of 4.5 to an increase of 3, with a median improvement of 1 (ordinal scale=10). The median score was 4.5 before and 5.5 after.

Confirmation of oesophageal intubation with direct laryngoscopy
Change in criterion 7 scores from the first to the second simulation ranged from a decrease of 1 (two residents) to an increase of 1 (three residents), with a median improvement of 0 \( (P=1.000 \text{ for the change}) \). In other words, eight of 14 SS subjects (57%) performed a second direct laryngoscopy in the first episode compared with nine of 14 (64%) in the second episode, with no significant difference between episodes.

Insertion of second tracheal tube
Three residents had changes in scores for question 8. One resident had a decrease of 1 and two residents had an increase of 1, and an overall median improvement of 0 \( (P=1.000 \text{ for the change}) \). In other words, one of 14 SS subjects (7%) inserted a second tracheal tube in the first episode, versus two of 14 (14%) in the second episode, without significant difference.
Comparisons between group SS and group JS
Second simulation total test scores were compared between the two groups using the exact Wilcoxon rank sum test. The median (minimum, maximum) total score was 5.5 (4, 7) for group SS and 5 (3, 9) for group JS. These group differences were not significant ($P=0.7758$).

Evaluation scores within each resident and test trial were averaged. Exact $\chi^2$ tests were then used to compare the two groups with respect to the second simulation outcomes for criteria 7 and 8. There were no group differences with respect to criteria 7 and $8$ ($P=0.6424$ and $P=0.5743$ respectively).

Prior visits, interim days, and seconds to ventilate
In the SS group, the mean number of prior visits to the simulation laboratory was 6.6 (SD 2.3) vs 6.8 (1.9) in the JS group. In the SS group, 208 (87) days elapsed between sessions vs 254 (88) in the JS group. Group SS required 134 s to ventilate in the first simulation vs 162 s in the second. Group JS required 239 s to ventilate in their second episode. There was no correlation between prior experience and criteria 7 and 8 scores and total scores. There was no correlation between either interim days or seconds to ventilate and the number of seconds to ventilate, but there was a trend towards a negative correlation between interim days and seconds to ventilate in the JS group ($r=0.714, P=0.0713$).

Pre-analysis questionnaire
Eighteen of 21 residents responded to the pre-analysis questionnaire (Table 2). There was unanimous agreement that medical simulation is advantageous to their education, that it provides a certain ‘shock value’ as an aid in remembering the teaching objectives, and that performance in simulation improves with simulation experience. A 2:1 majority felt that the same scenario should be repeated to reinforce new behaviours, while the same majority felt that they learn better in actual clinical settings. Whereas 94% had personally intubated the oesophagus, only 39% had ever confirmed it with a second direct laryngoscopy, and 33% said they had ever actually inserted a second tracheal tube, before their experience in the simulation laboratory.

Post-analysis questionnaire
Eleven of 21 residents responded to the post-analysis questionnaire (Table 3). There were a number of reasons why the residents did not follow the suggested management. The technique was generally viewed as new and out of the ordinary, and was not supported or reinforced by other teaching faculty. Some clinicians literally obstructed the resident from performing this technique in clinical situations. It was deemed unnecessary by several residents, particularly in routine situations, and might have been adopted more fully if there was some correlation to actual risk-prevention. Such correlation generated two testimonials. In one instance of oesophageal intubation, a former subject instructed the nurse anaesthetist to leave the first tube in place. The patient then vomited through the first tube as the second was successfully placed into the trachea. Another oesophageal intubation followed difficult and repeated attempts to visualize the glottis. The remaining tube then guided the tracheal tube into place. Trainees also requested practice, repetition and shorter intervals between episodes. Many could not recall practising the technique after the first scenario. However, the trainees unanimously felt that simulation is beneficial.

Discussion
This retrospective study was designed to document a learning outcome in the management of oesophageal intubation and to investigate the reasons for that outcome. Before these simulation sessions, the trainees had no formal instruction in the management of oesophageal intubation and were subjected to the individual teachings of the attending anaesthesiologists. Such was the impetus for this formal training in simulation. We therefore do not believe the residents were biased by other formal curricula.

Residents believe that simulation provides a certain shock value to stimulate learning. However, we could not simulate regurgitation of gastric contents, nor did we associate oesophageal intubation with a difficult airway. The oesophageal intubation itself had little shock value. Thus, the recommended procedure may have seemed unnecessary, offering no proof of benefit. The subsequent portion of each scenario may have made a greater impression on the trainee, since anaphylaxis and haemorrhagic shock are more dramatic, and this may have detracted from the recall of our suggestions for managing oesophageal intubation. In the two clinical testimonials, the clinicians were faced with the immediate and real consequences of oesophageal intubation. Vomiting through the oesophageal tube and localization of an otherwise hidden glottis were both dramatic episodes and provided enough shock value to convince these clinicians that they had learned a valuable lesson. This experimental learning is consistent with the trainees’ response in Table 2, Question 5, whereby they state learning is enhanced more in actual practice than in simulation. Our clinical practice norm did not include or support this method. Regardless, we demonstrated that a procedure first learned in simulation was in fact applied to the actual clinical setting, even though we could not demonstrate its application to another simulated session.

Assessment in simulation must be reliable and valid to demonstrate outcomes. Our assessment of the trainees’ performance is believed to be reliable, on the basis of the
exceptional inter-rater reliability of 100%. We did not, however, test for consistency between assessments, since only one assessment was made before (and only one after) the instructional intervention. Our assessment was valid for two defined reasons. Face validity is a measure of how appropriate the test items are to the purpose of the assessment, and the criteria in Table 1 are generally accepted performance measures in the management of oesophageal intubation.\(^8\) Criteria 7 and 8 have content validity because they directly measure and represent the skills of this instructional process. The ability to score technical criteria consistently, especially when the event occurs only once by a single individual, is also documented.\(^11\) \(^12\) Although there was a significant improvement in overall performance, with high reliability (concordance 0.829–1.000), we did not apply Cronbach’s test for internal consistency among these 10 evaluation criteria, since our primary objective was the new strategy of criteria 7 and 8.

Perhaps our failure to change management is attributed to a lack of construct and criterion validity. Apparently we failed to impress on the trainees that our management strategy was better than the conventional strategy of removing the first tube before placement of the second. Nor did we know whether the JS group should have performed better than SS. There are suggestions that an initial observational group performs better after receiving a debriefing than those who must first perform the scenario.\(^13\) However, our JS group was compared with the group (SS) that had both experience and debriefing. A recent review by Byrne and Greaves\(^10\) indicated that very few performance assessments in simulation were designed to investigate the validity or reliability of the assessment systems. Criterion validity relates the subject’s performance to established standards of practice. They found that most of the studies relied upon compliance with recognized, yet unvalidated treatment algorithms. We do not believe, now, that our performance criteria 7 and 8 had ‘criterion validity’ because they did not represent the standard of practice.

The Delphi procedure is a systematic group strategy that can be used to determine the most appropriate steps in the management of a critical incident.\(^14\) It relies upon roundtable expert evaluation, feedback, suggestion and modification, and provides validity for the proposed solution being taught or evaluated. When Forrest and colleagues\(^15\) determined the criteria of performance for rapid sequence induction by a modified Delphi technique, the ability of novice anaesthesia practitioners to improve their performance was statistically validated\(^16\) and demonstrated construct validity when compared with the performance of more experienced experts.

Experience in simulation must be frequent enough to reinforce the learning objectives. Data from Schwid and O’Donnell\(^17\) confirm that appropriate management of advanced cardiac life support scenarios is linked to short (<6 months) intervals between training sessions. Our interval of 7.7 months was too excessive to demonstrate retention of a procedure that was not practised, or reinforced clinically. We may also have reinforced this procedure inconsistently during the debriefing, which is a major limitation of retrospective research.

In summary, we were unable to show improvement in the specific technical management of oesophageal intubation after simulation training. These negative results reinforce the importance of thoroughly validating the assessment of performance and of providing the repetition and reinforcement needed to change behaviours.

References