British Journal of Anaesthesia **107** (4): 533–9 (2011)
Advance Access publication 9 June 2011 · doi:10.1093/bja/aer160

BJA

CLINICAL PRACTICE

Complex procedural skills are retained for a minimum of 1 yr after a single high-fidelity simulation training session†

S. Boet[1]*, B. C. R. Borges[1], V. N. Naik[2], L. W. Siu[1], N. Riem[1], D. Chandra[1], M. D. Bould[3] and H. S. Joo[1]

[1] Department of Anaesthesia, St Michael's Hospital, University of Toronto, 30 Bond Street, Toronto, ON, Canada M5B 1W8
[2] Department of Anaesthesia, The Ottawa Hospital, University of Ottawa, Ottawa, Ontario, Canada
[3] Department of Anaesthesia, The Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, Ontario, Canada

* Corresponding author. E-mail: sylvainboet@gmail.com

Editor's key points

- The use of simulation in education and training is well established.
- This study aimed to see how long the learning can be retained after a simulation training session.
- The complex procedural skills were found to be retained for a minimum of 1 yr.
- This study provides important information regarding how often the training should be repeated.

Background. Simulation has been shown to be effective in teaching complex emergency procedural skills. However, the retention of these skills for a period of up to 1 yr has not been studied. We aimed to investigate the 6 month and 1 yr retention of the complex procedural skill of cricothyroidotomy in attending anaesthetists using a high-fidelity-simulated cannot intubate, cannot ventilate (CICV) scenario.

Methods. Thirty-eight attending anaesthetists participated individually in a high-fidelity-simulated CICV scenario (pretest) that required a cricothyroidotomy for definitive airway management. Immediately after a debriefing and structured teaching session on cricothyroidotomy insertion, subjects managed a second identical CICV scenario (post-test). Each anaesthetist was randomized to either a '6 month retention' or a '12 month retention' group. No further teaching occurred. At their respective retention times, each anaesthetist managed a third identical CICV scenario (retention post-test). Two blinded experts independently rated videos of all performances in a random order, using a specific checklist (CL) score, a global-rating scale (GRS) score, and procedural time (PT).

Results. Subjects from both groups improved on their cricothyroidotomy skill performances from pretest to immediate post-test and from pretest to retention post-test, irrespective of the retention interval; CL mean (SD) 8.00 (2.39) *vs* 8.88 (1.53), $P=0.49$; GRS 28.00 (7.80) *vs* 31.25 (5.31), $P=0.25$; PT 102.83 (63.81) s *vs* 106.88 (36.68) s, $P=0.73$.

Conclusions. After a single simulation training session, improvements in cricothyroidotomy skills are retained for at least 1 yr. These findings suggest that high-fidelity simulation training, along with practice and feedback, can be used to maintain complex procedural skills for at least 1 yr.

Keywords: airway management; patient simulation; retention

Accepted for publication: 26 April 2011

Many emergency procedures are performed rarely but are of vital importance when needed. Examples of such procedures include needle thoracocentesis, cricothyroidotomy, and urgent thoracotomy for trauma.[1] These complex procedural skills must be learned and retained for potentially long periods of time before they are needed. The duration between acquisition of the skill and its eventual performance is referred to as the retention interval.[2] Skill decay has been shown to be problematic in this context, risking patient safety.[2 3] One of the main factors explaining decay of proficiency is the length of the retention interval.[2] Despite this potential decay, it is expected that a physician will perform the complex procedural tasks effectively in a clinical situation. Retention of skills after lectures, workshops, or courses has been demonstrated to be poor in several domains after only a few months.[4]

Although retention of complex, elective, and frequently performed procedural skills have been described,[5] the 1 yr retention of complex, rarely performed, and urgent procedural skills after a single session of high-fidelity simulation training has not been studied. We chose to study cricothyroidotomy, a procedure that has been validated for procedural skills testing,[6–8] as it may serve as a surrogate for other rare but emergently performed complex procedures.

†This article is accompanied by Editorial I.

Cricothyroidotomy is the final, rarely performed,[9] life-saving option in all cannot intubate, cannot ventilate (CICV) airway-management algorithms.[10 11] Consequently, anaesthetists have very little practical experience or confidence in performing this emergent intervention.[12 13] Since this procedure is performed only in crisis situations, it is crucial to optimize teaching and training methods for this skill outside the clinical realm.[14] The few studies investigating the retention of cricothyroidotomy skills have all used retention intervals of only ≤6 months.[15 16] In the absence of data regarding longer retention intervals, it has been suggested that formal simulation training should be provided at 6 month intervals by every anaesthesia department.[8 15 16] However, this short interval may be impractical.

This study examined the learning and 1 yr retention of emergency percutaneous cricothyroidotomy skills by attending anaesthetists during a high-fidelity-simulated CICV scenario. We hypothesized that a single simulation training session would improve the cricothyroidotomy performance both at 6 and 12 months for attending anaesthetists. We also hypothesized that retention in cricothyroidotomy performances would not be different between the 6 and 12 month retention groups after simulation training.

Methods

After Institutional Research Board approval (St Michael's Hospital, Toronto, Ontario, Canada), all 40 attending anaesthetists in a tertiary-care teaching hospital were offered the opportunity to participate. Thirty-eight subjects consented for enrolment. Informed consent was obtained in addition to a confidentiality agreement to prevent details pertaining to the clinical scenario from being disseminated before the end of the study. Subjects agreed to participate in two simulated scenarios with a third scenario at 'a later time, months later'. Subjects did not know the nature of any of the scenarios (CICV), nor the assessment criteria.

Study design and intervention

This was a single-blinded randomized study with repeated measures based on time. Subjects were block-randomized and assigned to the 6 or 12 month retention group according to the computer-generated allocation sequence of random numbers, allowing for allocation concealment and blinding to the debriefing instructor and raters. All subjects had a baseline measurement (pretest) of cricothyroidotomy performance followed by simulation teaching and then a post-test of cricothyroidotomy performance during a single session. The 1 yr effect of a single simulation training session was then studied by retesting the subjects either at 6 or 12 months based on their group allocation.

All scenarios were completed in a simulated operating theatre environment containing a high-fidelity manikin (SimMan®, Laerdal, Kent, UK), which has an anatomically accurate larynx that is designed for the practice of cricothyroidotomies, standard monitors (ECG, NIPB, Sp_{O_2}, E'_{CO_2}), and the anaesthesia machine (Datex® Corporation, St Laurent, Quebec, Canada). Before the studied simulation sessions, a standardized structured orientation session was held for each subject. Afterwards, each anaesthetist was familiarized with the simulator by participating in an introductory airway-management scenario that did not require cricothyroidotomy. Then, each subject participated individually in a high-fidelity-simulated anaesthesia crisis CICV scenario (pre-test). The subject was called to help with the intubation attempt already underway. On arrival, the simulated patient's oxygen saturation was 89% and decreased by 10% every minute. All methods of intubation were intended to be unsuccessful as the manikin was set up in the CICV configuration. No backup help was available. The scenario was designed to necessitate an emergency percutaneous cricothyroidotomy. The equipment available for the study was a 4.0 mm Melker kit (C-TCCS-400; Cook Inc., Bloomington, IN, USA). The scenario only ended with successful cricothyroidotomy performed by the subject, as defined by positive capnography.

After the first CICV scenario, personalized and video-assisted expert debriefing was provided. This included practical instructions on percutaneous cricothyroidotomy insertion, management guidelines of the emergency airway, and non-technical skills for crisis resource management. Subjects had practical 'hands-on' instructions detailing each step of percutaneous cricothyroidotomy insertion. They had the opportunity to practice on the same manikin but in a static environment (i.e. not in a crisis scenario and without alarms), with appropriate concurrent feedback from the instructor. Subjects then reviewed the video of the entire scenario. The whole teaching and debriefing session lasted for 1 h. The same individual conducted all debriefing sessions (L.W.S.). Subjects then managed an immediate post-test scenario. The subjects then returned for a retention post-test scenario at either 6 or 12 months after their training. The exact same CICV scenario was used in pretest, immediate post-test, and retention post-test. Subjects did not have prior knowledge of the content of any scenario.

Measurement instruments and outcomes

Subject characteristic data including gender, age, previous simulation exposure, previous cricothyroidotomy experience on both patients and manikins, and exposure (either manikin or clinical) to cricothyroidotomy between the immediate post-test and either retention post-test were collected (Supplementary Appendix 1).

Performance of the subjects was assessed with three parameters: a task-specific checklist (CL) adapted from Friedman and colleagues,[6] a global-rating scale (GRS), and procedural time (PT). Each step of the CL was scored either 0, 1, or 2, respectively, when it was not performed, poorly performed, or performed well (Supplementary Appendix 2). The GRS uses general descriptors and focuses on the overall procedural performance of the subject (Supplementary Appendix 3).[6] Friedman and colleagues[6] have demonstrated face and content validity and inter-rater reliability for both the CL and GRS. PT was defined as the time between when the subject first grasped any equipment from the cricothyroidotomy kit and the time of successful ventilation through the cricothyroidotomy.

Two evaluators with expertise in simulation were recruited and trained by the primary investigators to evaluate subjects using the CL and GRS. Training of the evaluators consisted of rating pre-recorded videotaped performances of simulated crises similar to the scenarios of this study using the CL and GRS. After data collection, the two evaluators independently reviewed and rated all videotapes in a random order. They were blinded to the subjects' randomization and test phase.

Statistical analysis

Sample size was calculated *a priori*. In the field of psychology and education, an effect size of >1.0 is considered large and acceptable for a given teaching intervention. Therefore, assuming an effect size of 1.0 (mean PT difference of 60 s, SD$=60$), $\alpha=0.05$, and a power of 0.8, we calculated that a total sample size of 34 subjects was required to test our hypotheses. Allowing for attrition, we invited 40 attending anaesthetists to participate in the study. Subject characteristic data were analysed using χ^2 test, Fisher's exact test, the Mann–Whitney test, and the unpaired *t*-test. The inter-rater reliability was assessed using the intra-class correlation coefficient for the total CL and GRS scores. Our outcome measures were analysed using a mixed analysis of variance (ANOVA). The 'outcome measure' (CL score, GRS score, PT) was treated as the dependent measure. The independent variables were the group allocation (6 month retention group, 12 month retention group) as the between-subjects factor and the test phase (pretest, immediate post-test, retention post-test) as the within-subjects factor. The mixed ANOVA was followed by planned comparisons using appropriate contrasts for multiple comparisons.[17] Although the individual Likert scales that make up the GRS scores are clearly ordinal in nature, these scores behave empirically as a parametric variable. Consequently, we, like other researchers, chose to use parametrical statistical analysis for our normally distributed data.[18] A two-tailed *P*-value of <0.05 was considered significant for all analyses. Statistical analysis was performed using SPSS 16.0 (SPSS Inc., Chicago, IL, USA).

Results

Thirty-four subjects completed the study (Fig. 1). Subjects' characteristics are summarized in Supplementary Appendix 1. None of the subjects had any exposure (neither manikin nor clinical) to cricothyroidotomy between the immediate post-test and the retention post-test.

The inter-rater reliability, measured by the intra-class correlation coefficient, was excellent: 0.947 ($P<0.001$) for the CL and 0.951 ($P<0.001$) for the GRS. The mean scores between the two experts were used for data analysis.

The main outcomes of the study are summarized in Figures 2–4 and Table 1. A mixed ANOVA detected a significant main effect of test phase (between pretest, immediate post-test, retention post-test) on the performance for the three outcome measures ($F_{2,60}=25.40$, $P<0.001$ on the CL; $F_{2,60}=38.39$, $P<0.001$ on the GRS; $F_{2,64}=8.42$, $P=0.001$ on the PT). No significant main effect of group (6 *vs* 12 month retention) on the performance for the three outcome measures was detected ($F_{1,30}=0.48$, $P=0.49$ on the CL; $F_{1,30}=1.37$, $P=0.25$ on the GRS; $F_{1,32}=0.12$, $P=0.73$ on the PT). The effect of group allocation (delay of retention) showed no interaction with respect to test phase (pretest, immediate post-test, or retention post-test) performance for the three outcome measures ($F_{2,60}=0.94$, $P=0.40$ on the CL; $F_{2,60}=0.58$, $P=0.56$ on the GRS; $F_{2,64}=0.18$, $P=0.84$ on the PT).

Planned comparisons revealed that subjects of both groups performed significantly better on their immediate post-test than on their pretest for the three outcome measures (CL, $P<0.001$; GRS, $P<0.001$; PT, $P=0.002$). As such, subjects of both groups performed significantly better on their retention post-test than on their pretest for the three outcome measures (CL, $P=0.001$; GRS, $P<0.001$; PT, $P=0.009$). Therefore, subjects improved on their performance for CL, GRS, and PT between pretest and immediate post-test and between pretest and retention post-test, irrespective of the delay of retention (group allocation).

The most frequently 'not performed' step in the CL at the pretest was 'Ventilation during cricothyroidotomy' ($n=13$ of 34) and it was also the most frequently 'not performed' step at the retention post-test ($n=9$ of 34). 'Correct use of dilator and cricothyroidotomy' was the item that most improved from pretest (correctly performed, $n=15$ of 34) to retention post-test (correctly performed, $n=30$ of 34). Finally, 'aspiration to identify trachea' improved from pretest (incorrectly performed, $n=10$ of 34) to retention post-test (incorrectly performed, $n=5$ of 34). The two other items of the CL were marginally incorrectly performed in all tests.

Discussion

This study showed that a single high-fidelity simulation cricothyroidotomy training session, including practice and feedback (debriefing), improved the procedural skills of attending anaesthetists and that this improvement was retained for at least 1 yr.

Previous literature showed decay in performance after 6 months and recommended that workshop training should be repeated every 6 months.[15 16] In contrast, our findings show that the complex procedural skill of cricothyroidotomy is retained for at least a year, when learned with high-fidelity simulation, along with practice and feedback. There are clear benefits to a yearly retraining or testing schedule. Studies have shown that lack of time is the first barrier to implement simulation on a regular basis for attending anesthesiologists.[19] Therefore, subjects are more likely to follow a yearly schedule when compared with a half- or quarter-year schedule.

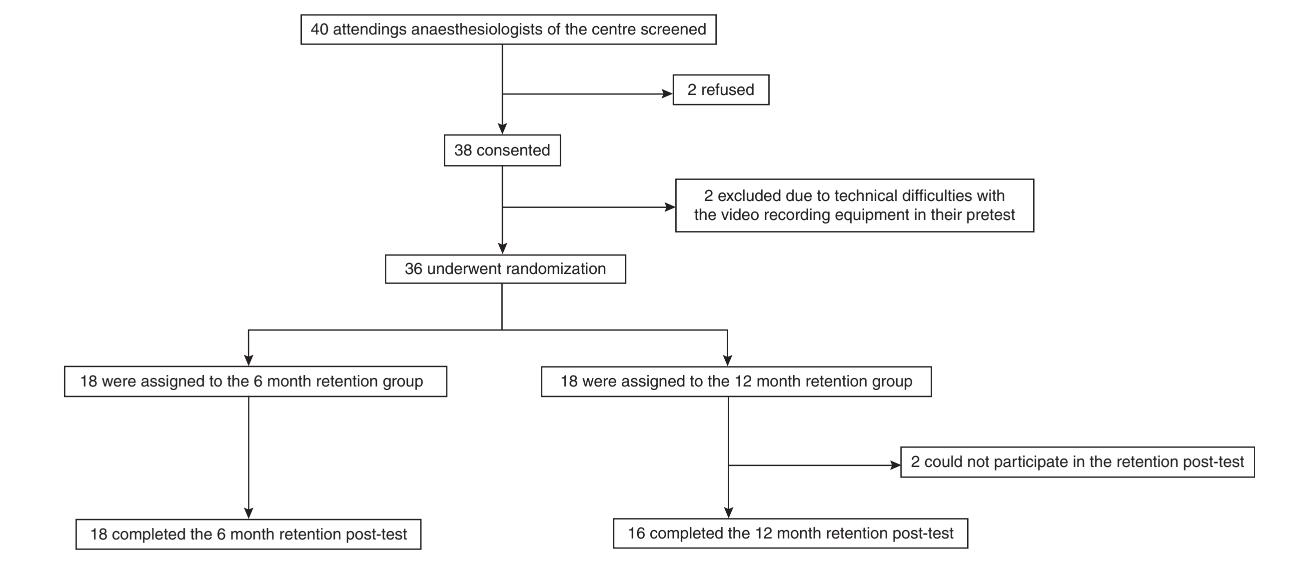

Fig 1 Screening, randomization, and completion of retention post-tests.

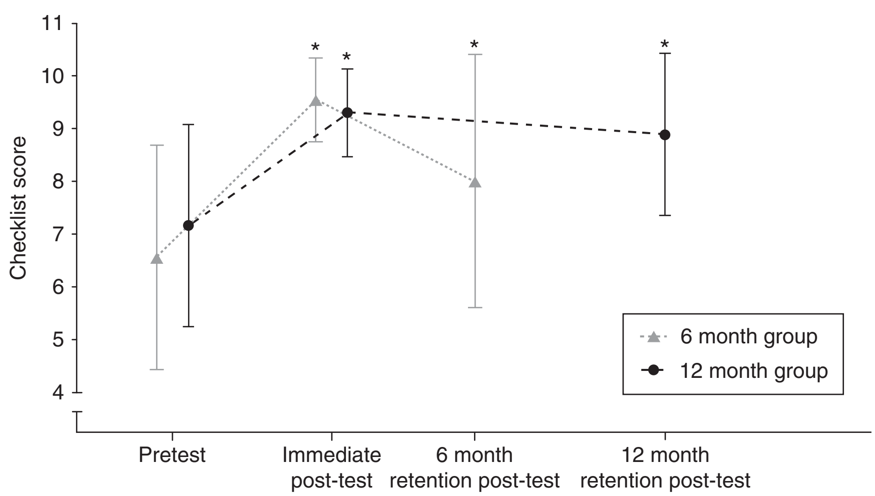

Fig 2 Performance in cricothyroidotomy skills on the CL score. Triangles and circles represent mean and error bars SD. *$P<0.05$ for comparison with pretest.

One potential reason to explain the longer retention period could be the realism of the scenario used to teach and test the performance in our study. In contrast to common teaching sessions composed of workshops, lectures, or low-fidelity static models, we chose to use a high-fidelity manikin associated with a dynamic environment and high-fidelity monitors in order to maximally reproduce the realism of CICV situations. We did not measure stress in our subjects; however, it is likely that the realism of high-fidelity simulation with alarms from the monitors induced a certain level of stress for the subjects. Research in psychology and psychoneuroendocrinology has concluded that the effect of stress on performance is determined by the cognitive appraisal of the situation by the subject under stress.[20] These effects appear to be the result of the individual's perception of the demands and resources of a situation. When an individual perceives a high level of resources (personal or environmental) to meet the demand (i.e. manage a CICV scenario), then the situation is considered as a challenge and a positive psychological state leads to an enhancement of performance.[21] When the demands are assessed as outweighing the resources of the individual, the situation is assessed as a threat and a negative psychological state leads to a decreased level of performance.[21] The realism and dynamic environment of the scenarios used in this study, when compared with the usual static models of

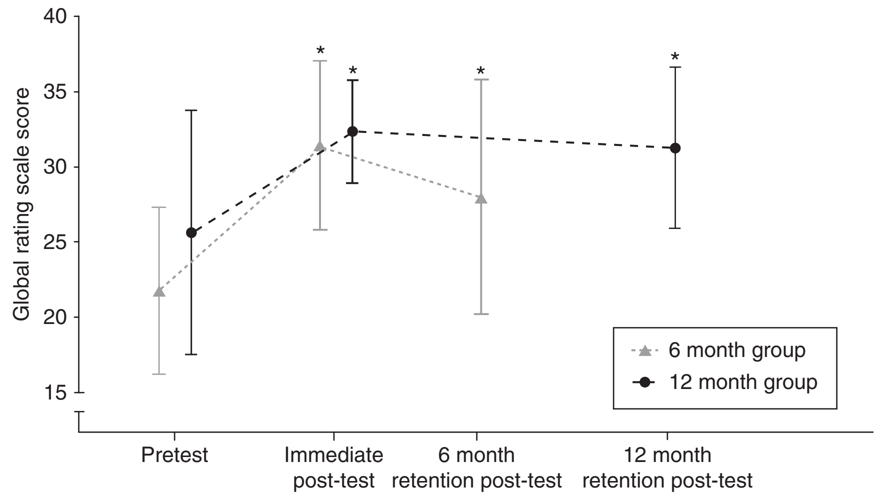

Fig 3 Performance in cricothyroidotomy skills on the GRS score. Triangles and circles represent mean and error bars SD. *$P<0.05$ for comparison with pretest.

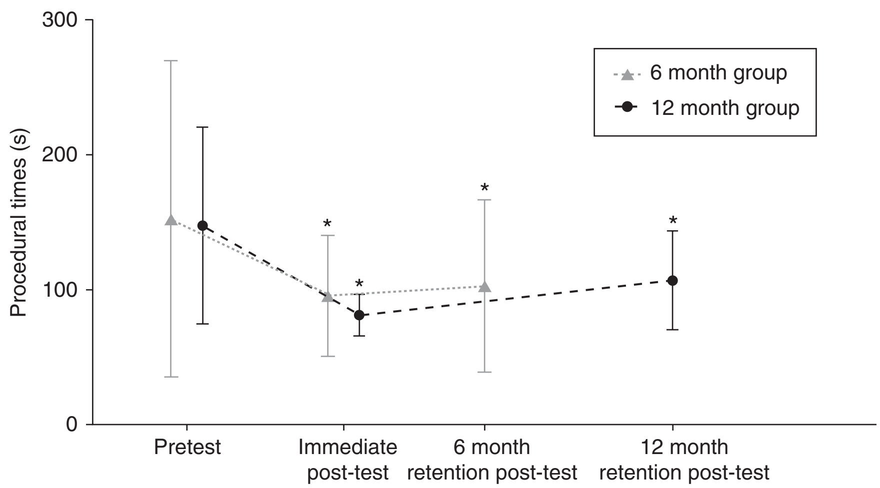

Fig 4 Performance in cricothyroidotomy skills on PT. Triangles and circles represent mean and error bars SD. *$P<0.05$ for comparison with pretest.

Table 1 Performance scores for all three tests phases. Mean (SD). *Significantly improved scores compared with pretest ($P<0.05$)

	Pretest		Immediate post-test		Retention post-test	
	6 months	12 months	6 months	12 months	6 months	12 months
Checklist (0–10)	6.56 (2.13)	7.16 (1.91)	9.53 (0.79)*	9.29 (0.83)*	8.00 (2.39)*	8.88 (1.53)*
Global-rating scale (0–35)	21.78 (5.54)	25.63 (8.10)	31.44 (5.63)*	32.36 (3.43)*	28.00 (7.80)*	31.25 (5.31)*
Procedural time (s)	152.61 (117.34)	147.56 (73.11)	95.50 (44.76)*	81.13 (15.47)*	102.83 (63.81)*	106.88 (36.68)*

training, may have improved performance if the subjects felt stressed and were able to appraise the situation as a challenge after training and not as a threat. This is only a hypothesis and of course further studies are needed.

The improved retention of skills found in our study compared with previous research may be partly explained by the episodic memory theory.[22] Remembering requires the two successive steps of encoding and retrieval. Encoding starts with the perception of an episode and finishes with memory trace. In our study, the episode characterized by the context of learning (practice in high-fidelity scenario and follow-up feedback) was likely to be memorable by attending anaesthetists. The training session included a practice (pretest) followed by feedback (debriefing), followed by another practice (immediate post-test). This structured training potentially induced a testing effect on the subjects. The testing effect refers to enhanced learning resulting from the act of retrieving information, when compared with simply reading or hearing the information. Karpicke and Roediger[23] recently demonstrated the critical role of retrieval practice in consolidating learning, specifically when repeated. Retrieval needs appropriate cues to occur.[22] Retrieval is improved when the context of retrieval is similar to the context of the encoding.[2] In our study, the necessary cue for retrieval could be the context of learning (scenario and emotional status of the subject). Retrieval during high-fidelity simulation with a dynamic scenario may have clinical relevance because it was as close to a real clinical emergency as possible. Moreover, feedback is considered to be crucial to learning from tests.[24] Additional feedback has been shown to further enhance the mnemonic benefits of testing.[25] Therefore, debriefing in our offered training may also partially explain the 1 yr retention of cricothyroidotomy skills. However, we cannot determine from our study which element of the practice, retrieval, or feedback has the most importance for potential improvement of episodic memory.

One of the strengths of our study may have been the recruitment of fully trained attending anaesthetists as subjects. In the past, the recruitment of residents for longitudinal studies may have been associated with contamination as the residents are generally taught on a routine basis as part of their training.[5 6] By using fully trained attending anaesthetists, we were able to avoid confounding variables and our results may be more relevant and applicable to concepts such as continuing medical education and recertification.

Time to perform a cricothyroidotomy in our study has been found to be longer than previously reported.[7 8 16] This is in keeping with John and colleagues[7] who demonstrated that time to perform a cricothyroidotomy on a high-fidelity simulator in a dynamic environment is likely to be longer than on a static low-fidelity simulator. One more possible reason to explain this difference is the type of technique and cricothyroidotomy kit used among studies. Others used a jet ventilation kit or a pre-assembled Melker CT set,[7 8 16] whereas we used the Melker emergency cricothyroidotomy catheter set that requires more steps.

Our study has several limitations. First, we chose to use the specific technique of inserting a Melker kit. Results might have been different with other cricothyroidotomy kits or other techniques. However, it seems likely that high-fidelity simulation training would be similarly useful for teaching other airway-management techniques. A further limitation is that we used simulation to answer our research question. Although retention during actual clinical practice is the fundamental question, human-controlled trials are difficult, if not impossible, to perform given the emergent nature and relative rarity of the procedure. We acknowledge that performance of cricothyroidotomy on manikins is artificial and different from real-life situations. However, literature seems to indicate transfer of skills from simulators to real patients for both non-technical and procedural surgical skills.[26 27] Another limitation is that we did not aim to teach the subjects to their maximal plateau of cricothyroidotomy skills before testing them. However, we were not trying to test retention of skills by 'experts'. We merely aimed to test the decay of learned skills after a single simulation-based education session.

In conclusion, our study demonstrated that a single high-fidelity simulation session, along with practice and feedback, improved the performance of attending anaesthetists with retention for at least 1 yr. These findings suggest that complex procedural skills can be maintained on a yearly basis after a high-fidelity simulation practice and feedback.

Supplementary material

Supplementary material is available at *British Journal of Anaesthesia* online.

Acknowledgements

The authors would like to thank the attending anaesthetists of St Michael's Hospital (Toronto, Ontario, Canada) for their participation in this study, Heinz Bruppacher for rating videos, Roger Chow and Paula Iantomasi for their technical support during simulation sessions, Rosane Nisenbaum for her statistical expertise, and Dr G. Hare for reviewing the manuscript.

Funding

Supported by the Department of Anaesthesia of St Michael's Hospital, University of Toronto (Toronto, Ontario, Canada).

References

1 Ahmad T, Ahmed S, Hussain N, Khan M. Role of urgent thoracotomy in improving the survival of patients with severe chest trauma. *J Coll Physicians Surg Pak* 2009; **19**: 575

2 Arthur W, Bennett W, Stanush PL, McNelly TL. Factors that influence skill decay and retention: a quantitative review and analysis. *Hum Perform* 1998; **11**: 57–101

3 Wisher R, Sabol M, Ellis J, Ellis K. *Staying sharp: retention of military knowledge and skills*. Report of the US Army Research Institute, July 1999; 1–32. Available from http://handle.dtic.mil/100.2/ADA366825 (accessed 14 May 2011)

4 Semeraro F, Signore L, Cerchiari EL. Retention of CPR performance in anaesthetists. *Resuscitation* 2006; **68**: 101–8

5 Castellvi AO, Hollett LA, Minhajuddin A, Hogg DC, Tesfay ST, Scott DJ. Maintaining proficiency after fundamentals of laparoscopic surgery training: a 1-year analysis of skill retention for surgery residents. *Surgery* 2009; **146**: 387–93

6 Friedman Z, You-Ten KE, Bould MD, Naik V. Teaching lifesaving procedures: the impact of model fidelity on acquisition and transfer of cricothyrotomy skills to performance on cadavers. *Anesth Analg* 2008; **107**: 1663–9

7 John B, Suri I, Hillermann C, Mendonca C. Comparison of cricothyroidotomy on manikin vs. simulator: a randomised cross-over study. *Anaesthesia* 2007; **62**: 1029

8 Wong DT, Prabhu AJ, Coloma M, Imasogie N, Chung FF. What is the minimum training required for successful cricothyroidotomy. *Anesthesiology* 2003; **98**: 349–53

9 Sagarin MJ, Barton ED, Chng YM, Walls RM. Airway management by US and Canadian emergency medicine residents: a multicenter analysis of more than 6,000 endotracheal intubation attempts. *Ann Emerg Med* 2005; **46**: 328–36

10 Practice guidelines for management of the difficult airway: an updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. *Anesthesiology* 2003; **98**: 1269–77

11 Henderson J, Popat M, Latto I, Pearce A. Difficult Airway Society guidelines for management of the unanticipated difficult intubation. *Anaesthesia* 2004; **59**: 675–94

12 Wong DT, Lai K, Chung FF, Ho RY. Cannot intubate–cannot ventilate and difficult intubation strategies: results of a Canadian national survey. *Anesth Analg* 2005; **100**: 1439–46

13 McNarry AF, Dovell T, Dancey FM, Pead ME. Perception of training needs and opportunities in advanced airway skills: a survey of British and Irish trainees. *Eur J Anaesthesiol* 2007; **24**: 498–504

14 Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med* 2003; **78**: 783–8

15 Kuduvalli PM, Jervis A, Tighe SQ, Robin NM. Unanticipated difficult airway management in anaesthetised patients: a prospective study of the effect of mannequin training on management strategies and skill retention. *Anaesthesia* 2008; **63**: 364–9

16 Prabhu A, Correa R, Wong D, McGuire G, Chung F. What is the optimal training interval for a cricothyroidotomy (CT)? *Can J Anesth* 2001; **48**: 59

17 Ruxton G, Beauchamp G. Time for some a priori thinking about post hoc testing. *Behav Ecol* 2008; **19**: 690

18 Carifio J, Perla R. Resolving the 50-year debate around using and misusing Likert scales. *Med Educ* 2008; **42**: 1150–52

19 Savoldelli GL, Naik VN, Hamstra SJ, Morgan PJ. Barriers to use of simulation-based education. *Can J Anaesth* 2005; **52**: 944–50

20 LeBlanc V. The effects of acute stress on performance: implications for health professions education. *Acad Med* 2009; **84**: S25

21 Tomaka J, Blascovich J, Kelsey R, Leitten C. Subjective, physiological, and behavioral effects of threat and challenge appraisal. *J Pers Soc Psychol* 1993; **65**: 248–60

22 Tulving E. Precis of elements of episodic memory. *Behav Brain Sci* 1984; **7**: 223–38

23 Karpicke JD, Roediger HL 3rd. The critical importance of retrieval for learning. *Science* 2008; **319**: 966–8

24 Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. *Med Educ* 2008; **42**: 959–66

25 Kang S, McDermott K, Roediger H III. Test format and corrective feedback modify the effect of testing on long-term retention. *Eur J Cogn Psychol* 2007; **19**: 528–58

26 Bruppacher HR, Alam SK, LeBlanc VR, *et al.* Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *Anesthesiology* 2010; **112**: 985–92

27 Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. *Ann Surg* 2008; **248**: 166–79