Evaluating the ORSIM® simulator for assessment of anaesthetists’ skills in flexible bronchoscopy: aspects of validity and reliability

P. A. Baker¹,², J. M. Weller¹,², M. J. Baker¹, G. L. Hounsell³, J. Scott²,³, P. J. Gardiner² and J. M. D. Thompson¹

¹University of Auckland, Auckland, New Zealand, ²Auckland City Hospital, Auckland, New Zealand, and ³Middlemore Hospital, Auckland, New Zealand

Abstract

Background: Developing expertise in flexible bronchoscopy is limited by inadequate opportunities to train on difficult airways. The new ORSIM bronchoscopy simulator aims to address this by creating virtual patients with difficult airways. This study aims to provide evidence on the validity and reliability of the ORSIM for assessment of subjects on both normal and abnormal airway simulations.

Methods: Novice, trainee, and expert subjects performed seven simulations of varying difficulty and scored the perceived difficulty for each. Time to completion was measured. Three blinded raters independently scored videos of each subject’s performance. We measured inter-rater agreement and the difference in raters’ scores between subject groups.

Results: We recruited 28 study subjects, generating 196 videos for analysis. Expert subjects consistently completed the scenarios faster than novices. Overall performance scores showed significant differences between subject groups (P<0.0001). Inter-rater reliability of scores was >0.8.

Conclusions: Our results provide initial evidence on the validity and reliability of the ORSIM bronchoscopy simulator, supporting its potential value in training and assessment.

Key words: airway management; computer simulation; education, continuing; scoring methods

Opportunities for flexible bronchoscopy in the operating room have decreased over the last decade with the introduction of videolaryngoscopes and supraglottic airways.¹ This is likely to have a deleterious effect on the development and maintenance of flexible bronchoscopy skills for tracheal intubation.

To address this problem, we have developed a new virtual reality simulator, the ORSIM® bronchoscopy simulator (Airway Simulation Limited, Auckland, New Zealand), which aims to address some of the issues of advanced airway training by incorporating virtual patients with difficult airways.² The ORSIM consists of hardware and software components that interact to create a high-fidelity virtual reality simulation (Fig. 1). A replica video bronchoscope is advanced by the user through a desktop sensor, which is connected to a dedicated laptop computer. The laptop software program includes multiple upper and lower virtual airways of varying complexity. Virtual oral, nasal, and conduit entry points can be selected, and the program provides recording, feedback, measurement, and relevant clinical data. This simulator has
A new bronchoscopy simulator (ORSIM) may be useful for developing expertise in fibre-optic bronchoscopy, but its validity and reliability are not known.

Virtual reality simulators can supplement clinical exposure and provide independent and objective assessment of performance in an environment that is safe for patients, learners, and teachers. Simulators designed for training and assessment of procedural airway skills should provide a valid representation of clinical patients and a reliable measure of performance across a range of airway problems and degrees of difficulty. Ideally, we would seek evidence that performance in a simulator predicts performance in the clinical environment; however, the scheduling of flexible bronchoscopy for patients with a suitable range of difficult airways is logistically and ethically difficult. As a first step, we set out to find evidence of validity and reliability of the ORSIM simulator when used for assessment of flexible bronchoscopy skills.

Our specific research questions were as follows.

(i) Timing: do expert subjects complete the simulations faster than trainees and novices?
(ii) Inter-rater reliability: do independent, blinded, external scorers agree on the standard of performance of subjects across a range of upper airway simulations designed to be of varying difficulty?
(iii) Construct validity: are the scores given by scorers for ORSIM performance significantly different for novices, trainees, and experts, and are the scores consistent with expected performance where:
   (a) Level of expertise is determined from subject self-report of numbers of flexible bronchoscopies?
   (b) Level of expertise is determined from the median of categorical scores for performance across seven simulations by the three scorers?

**Methods**

This study was conducted at Auckland City Hospital and the University of Auckland Clinical Skills Centres, Auckland, New Zealand (from February 2012 to July 2012); it was approved by the Northern X Regional Ethics Committee (approval number NTX/12/EXP/037); and written informed consent was gained from subjects. All data were de-identified before analysis.

**Subjects and sampling**

We sought subjects who had no prior experience with the ORSIM bronchoscopy simulator and who satisfied the following criteria: an expert group comprising anaesthetists who had performed more than 50 flexible bronchoscope intubations in their career; a trainee group comprising anaesthetic registrars who had performed one to 10 flexible bronchoscope intubations in their career; and a novice group comprising anaesthetic technicians who had never performed a flexible bronchoscope intubation before, but were familiar with the equipment and the procedure.

We recruited a volunteer sample of subjects: anaesthetists and trainees attending other training courses; and anaesthetic technicians from Auckland City Hospital. We collected subject characteristic data on their level of training, years of practice, and lifetime and 12 month experience with flexible bronchoscopy.

**Simulations**

Before commencing the study, the subjects were given scripted verbal and written instructions about how to use the simulator and then watched an instructional video. They received no further tuition throughout the study.

Each subject performed the same seven scenarios in the same order. The scenarios were as follows: (1) a normal oral intubation; (2) a normal nasal intubation; (3) a tumour partly obstructing the larynx; (4) a tumour on the base of the tongue with very limited laryngeal access; (5) a retropharyngeal abscess; (6) a narrow laryngeal airway because of an obstructing lesion arising from the vestibular fold; and (7) an inflamed, swollen epiglottis. The exercises required the operator to pass the model flexible bronchoscope through the virtual airway, avoiding airway trauma from collisions, while being timed from the point of entry, through the upper airway to the main carina, where the scenario ended. Each scenario was recorded on the laptop computer.
Scorer training

Three independent senior anaesthetists, not involved in the study days, were selected as scorers on the basis of their clinical experience and their experience in teaching flexible bronchoscopy, both in clinical situations and in airway courses. Initial scorer training consisted of a meeting to discuss assessment criteria and scoring, using three sample videos from each group. The scorers then independently rated 30 randomly selected videos. They discussed their scoring at a subsequent training session, and agreed to use the following criteria to guide their scoring: economy of movement; stability of image; maintenance of ideal flight path; avoidance of collisions; purposeful movement; and speed.

Data collection

The time to completion of the simulation was downloaded from the ORSIM computer.

The three scorers independently assessed all videos from each of the subjects, in random order and blinded to subject identity. Each ORSIM video was scored using two scales: a 100 mm visual analog scale (VAS), with left and right limits of the line labelled ‘novice’ and ‘expert’, respectively; and a three-point categorical scale, labelled novice, trainee, and expert.

Analyses

The selection of subjects for the study was opportunistic, being from training workshops. The study was planned to have 30 subjects, with 10 in each group and thus 70 scored scenarios in each group. There was no previous information available for the VAS, but this did not reach statistical significance (P=0.70; Table 1).

Analyses

The selection of subjects for the study was opportunistic, being from training workshops. The study was planned to have 30 subjects, with 10 in each group and thus 70 scored scenarios in each group. There was no previous information available for the VAS, but this did not reach statistical significance (P=0.70; Table 1).

Timing

Novices were significantly slower than experts by the mean difference of 78 s (P<0.0023); trainees were 10 s slower than experts, respectively: and a three-point categorical scale, labelled novice, trainee, and expert.

Table 1 Duration of anaesthesia career, mean self-rated expertise score using a 100 mm visual analog scale, and mean scenario completion times recorded in seconds (sd) by the simulator

<table>
<thead>
<tr>
<th>Subject</th>
<th>Years of practice</th>
<th>Fibre-optic bronchosopies in career</th>
<th>Fibre-optic bronchosopies in 12 months</th>
<th>Subjective expertise score [mm; mean (sd)]</th>
<th>Objective scenario time [s; mean (sd)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert</td>
<td>7–37</td>
<td>&gt;50</td>
<td>1–20</td>
<td>72 (1.1)</td>
<td>116 (17)</td>
</tr>
<tr>
<td>Trainee</td>
<td>2–4</td>
<td>1–10</td>
<td>1–5</td>
<td>24 (1.7)</td>
<td>126 (26)</td>
</tr>
<tr>
<td>Novice</td>
<td>Nil</td>
<td>0</td>
<td>0</td>
<td>2 (1.5)</td>
<td>195 (17)</td>
</tr>
</tbody>
</table>

Table 2 Participant self-assessment of difficulty for each scenario was measured in millimetres on a visual analog scale and recorded as means (sd), with the left and right ends of the 100 mm line labelled ‘easy’ and ‘difficult’, respectively

<table>
<thead>
<tr>
<th>Subject</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
<th>Scenario 5</th>
<th>Scenario 6</th>
<th>Scenario 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert</td>
<td>17 (19.8)</td>
<td>24 (17.6)</td>
<td>19 (17)</td>
<td>81 (21.1)</td>
<td>52 (21.1)</td>
<td>45 (25.5)</td>
<td>58 (35.6)</td>
</tr>
<tr>
<td>Trainee</td>
<td>25 (19.6)</td>
<td>39 (18.3)</td>
<td>36 (17)</td>
<td>84 (16.5)</td>
<td>66 (19.4)</td>
<td>60 (14.9)</td>
<td>92 (6.4)</td>
</tr>
<tr>
<td>Novice</td>
<td>52 (24.3)</td>
<td>56 (28.9)</td>
<td>43 (24.5)</td>
<td>95 (5.3)</td>
<td>83 (18)</td>
<td>57 (22.8)</td>
<td>92 (11.5)</td>
</tr>
</tbody>
</table>
of previous flexible bronchoscopies performed. Figure 2a illustrates the second method, where VAS scores show the level of expertise derived from the median of categorical scores by the three scorers for performance across seven simulations.

The agreement between the subjects’ self-reported level of experience and the determination by the scorers was r=0.54. Of the 12 anaesthetic technicians, seven were assessed as novices and five as trainees. All trainees were correctly assessed as trainees. Of the nine defined as experts, five were assessed as performing at the level of an expert and four at the level of a trainee.

Discussion

In this study, we have provided some evidence supporting the reliability and validity of the new ORSIM bronchoscopy simulator for training and assessment. Expert subjects consistently completed the scenarios faster than novices. Inter-rater reliability of VAS scores by three blinded scorers was >0.8. Scores showed significant differences (all P<0.0001) between each of the subject groups (novice, trainee, and expert), with progressively higher scores from novice to expert. This supports the construct validity of the ORSIM simulator for assessment of flexible bronchoscopy.

Previous studies of tracheal intubation skills using a flexible bronchoscope have tested subjects only on normal airways, or airways comparable to manikins with normal airways, after training on models, manikins, or virtual reality simulators.5–9 Our study has extended the field by incorporating simulated abnormal airways. Using normal airways for training and assessment limits the discriminative ability of the assessment by creating a low ceiling effect. Furthermore, the clinical need for expertise in flexible bronchoscopy in tracheal intubation is in the anatomically difficult airway, rather than the normal airway.

A number of validation studies have failed to demonstrate differences in performance between groups with different levels of expertise in flexible bronchoscopy. This has been attributed to several factors, including testing with tasks that are too easy, testing with too few procedures, and failing to use performance indicators with clinical relevance to supplement the simulator metrics.10–12 One study of bronchoscopy performance, using the Accutouch (CAE Healthcare, Montreal, Québec, Canada), was able to differentiate between novice, trainee, and expert subjects by using the following: exercises of varying difficulty; multiple test items; a standardized setting; and a scoring system that was clinically relevant.12 The Accutouch was an early virtual reality bronchoscopy simulator, which included bronchoscopy, gastroscopy, and colonoscopy. In our study, we used a similar approach, but with a simulator designed only for bronchoscopy, which includes a range of difficult upper airway navigations. We used multiple tests of varying difficulty, clinically relevant scenarios, a standardized setting, and a clinically relevant scoring system. Future validation studies should take these factors into account.

We have also demonstrated a difference between the judgments of experts on the level of performance and the self-reported experience of subjects with flexible bronchoscopy. This suggests that previous numbers of patients may not be a reliable measure of expertise in managing patients with anatomically challenging airways.

The issue of distinguishing between the three groups using the VAS score was considered. We think that the important groups to differentiate are trainee vs expert. A valid and reliable score could then be used to determine when a trainee is suitably skilled on the simulator to be able to perform at the standard of an expert on a patient with a difficult airway. Our results suggest that a VAS score of 70 or greater only identifies experts (Fig. 2a and b).

Limitations

In this study, we have not explored transfer of skills in difficult airway simulations to performance in flexible bronchoscopic intubations of patients with difficult airways in the clinical

Table 3 Average correlations across all scenarios and pairs of scorers for the overall score and the various component scores

<table>
<thead>
<tr>
<th>Score categories</th>
<th>Overall</th>
<th>Movement</th>
<th>Stability</th>
<th>Flight</th>
<th>Collisions</th>
<th>Decisions</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>71</td>
<td>45</td>
<td>40</td>
<td>69</td>
<td>55</td>
<td>65</td>
<td>58</td>
</tr>
</tbody>
</table>
environment. Previous flexible bronchoscopy simulator studies have established this relationship between performance in training on the simulator and subsequent clinical performance in patients with normal airways.\textsuperscript{5, 11} It would be ethically and logistically difficult to undertake such a study with anatomically difficult or abnormal airways.

The characteristics of the simulator and the blinded nature of scoring with videos limited our ability to measure some important aspects of flexible bronchoscope use, such as patient preparation, positioning, and appropriate handling of the bronchoscope. In a training environment, these skills could be taught in conjunction with the simulator.

In this study, we chose anaesthetic technicians as our novice group because of their theoretical knowledge of airway anatomy, pathology, and equipment. They were, however, novices because they had no experience at performing flexible bronchoscopy. The technicians were also chosen because of the wide variability and technical experience seen in our trainee registrars. We acknowledge that the novice technician group does not represent a clinically transferrable study group.

Future directions
This study relied on external scoring of large numbers of videos by independent scorers. This was a time-consuming exercise and probably not feasible for practical use in research or assessment. Future studies will focus on intrinsic automated scoring with computer-generated metrics. The scores for performance generated by external expert scorers provide a basis for establishing the validity of computer-generated metrics. Randomization would allow us to differentiate between learning experience and differences in scores between scenarios. Testing computer-generated metrics against expert scores, using a random sample of participants, could be a useful next step.

Conclusion
This study supports the potential use of the ORSIM bronchoscopy simulator as an assessment tool to measure procedural skill with a flexible bronchoscope. Although simulation has provided a useful supplement to clinical experience, it has been previously restricted to training and assessment on normal airways, requiring the practitioner to return to the operating room to gain the skill and experience to manage difficult airways. There is now the opportunity for simulation to accelerate training on difficult airways and to assess flexible bronchoscopy performance at the higher level of expertise required for management of complex airway problems.

Authors’ contributions
Study design/planning: P.A.B., J.M.W., J.M.D.T.
Study conduct: P.A.B., M.J.B., G.L.H., J.S., P.J.G.
Data analysis: J.M.W., J.M.D.T.
Writing the manuscript: P.A.B., J.M.W., M.J.B., G.L.H., J.S., J.M.D.T.
Revision of the manuscript: P.A.B., J.M.W., M.J.B., G.L.H., J.S., P.J.G., J.M.D.T.

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

Acknowledgements
The ORSIM was supplied free of charge for this study by Airway Simulation Limited, who manufacture the ORSIM. We are grateful to C. Baird, who assisted with data entry. We also acknowledge Professor Alan Merry, Head of the School of Medicine, Faculty of Medical and Health Science, University of Auckland, Auckland, New Zealand, and Dr Mohsen Davoudi, Chief of University of California Irvine Division of Pulmonary Diseases and Critical Care Medicine, University of California, Irvine, CA, USA, who reviewed this manuscript and offered valuable advice.

Declaration of interest
P.A.B. and M.J.B. retain a commercial interest in the ORSIM bronchoscopy simulator and have been involved in the design and development of this product. They are the owners of Airway Simulation Limited, which manufactures the ORSIM. In this study, they were excluded from data entry, scoring, or data analysis. Their commercial involvement in the ORSIM was made clear to the other authors and the subjects at the commencement of the study. The authors who were involved in scoring and data analysis for this study have no commercial interest in the ORSIM. J.M.W. serves on the Editorial Board of the British Journal of Anaesthesia. No other conflicts of interest are declared by the authors.

Funding
Institutional and departmental sources.

References

Handling editor: T. Asai