Inherent variability in airway characteristics of simulation manikins: is it time we standardised assessments of crisis management skills?

Balakrishnan Ashokka,1 Krishnasamy Narendiran,2 Abhijit Bhattacharya,3 Dinker Pai,4 Shen Liang,5,6 Shoba Subramanian,2 Ernest T Larmie,3 Fun Gee Chen1

ABSTRACT

Introduction Learning of simulation-based crisis management skills involves technologically advanced manikins and use of automated scenarios. Progressions in preprogrammed scenarios require finite task completion such as successful airway intubations for achieving optimal learning outcomes aligned to curricular goals. The study was set to explore the existing variability among various simulation manikins in use at our institute for undergraduate medical education.

Methods 56 final-year undergraduate students, who had received prior training in airway management skills, performed intubations on each of the 5 different manikins (56 x 5=280 intubations). The manikins used were the Human Patient Simulator (HPS), iStan & Emergency Care Simulator (ECS) from CAE Healthcare and Mega Code Kelly (MCK) and Airway Trainer (AWTR) from Laerdal. The students’ performances were compared for success rates, ease of intubation, grade of laryngeal visualisation and presence of tooth injury on the manikins. Data from the intubations were cross-tabulated and evaluated by general estimating equation analysis using the Poisson model.

Results iStan had the higher rates of failure to intubate (64.3%). iStan (62.5%) and HPS (57.1%) had statistically significant teeth injury (p<0.0001) compared to other manikins. HPS and AWTR had the least difficult grades of laryngeal visualisation (Cormack Lehane grades 1 and 2), while the most difficult grade of visualisation (Cormack Lehane grades 3 and 4) was reported in ECS (44.6%).

Conclusions Each of the high-technology manikins used in automated scenarios for crisis management teaching and learning has heterogeneity in airway features. Since frequent airway management is a critical component of simulation scenarios, this can affect student performance when these manikins are used for formative and summative high-stakes assessments.

INTRODUCTION

Simulation-based crisis management education and training have been proven to be effective in acquisition of technical and non-technical skills relevant to the management of critical events.1-3 Healthcare applications of simulation have focused on airway management as an essential skill required for medical students, residents and all other healthcare professionals. Learning this skill needs adequate time. Practice with task trainers and advanced simulators is known to improve correct handling of simulated clinical conditions.4 Advanced airway skills training including correct mask-holding techniques, insertion of supraglottic devices, etc, require the use of advanced anatomically appropriate simulators.5 Given the alarming rates of iatrogenic morbidity and mortality around the world stemming from medical errors, it is necessary to minimise these by providing adequate simulation-based training. Also, the concern about the ability of inexperienced trainees to handle emergency medical situations under highly stressful conditions in vulnerable patients has driven the use of simulators in crisis management training.6-7

Simulation faculty use automated preprogrammed scenarios in technologically advanced manikins that have definitive end points (eg, intubation, needle thoracocentesis) to transit to the next clinical stage of the simulation session.8 Even though advanced manikins are not meant to be used as task trainers, variations in airway intubation difficulty in these manikins can directly affect the preprogrammed ‘patient’ outcomes (ie, scenario progressions) during these sessions.9 We therefore explored the variability in airway characteristics in the manikins that were being used at our institute in undergraduate training and assessments.

METHODOLOGY

At the Asian Institute of Medicine, education and Technology (AIMST) University Malaysia, undergraduate medical students were taught through a simulation-based integrated teaching programme. In their second and third years, basic airway management (basic life support, bag-mask ventilation, endotracheal intubation) was taught through component-task trainers such as the Airway Trainer (AWTR) (Laerdal, Stavanger, Norway). The teaching was complemented by formal lectures and web-based videos on component airway skills. Students were taught about the significance of the alignment of the three airway axes—oral, pharyngeal and laryngeal. The students in their fourth year were taught advanced airway management skills including airway assessment and management of difficult situations, advanced cardiac-life support skills through high-technology mannikin-based scenarios. Team-based advanced clinical scenarios were used to teach crisis management in the final year of undergraduate education.

For the study, 56 final-year MBBS students (n=56) who had received prior simulation-based component airway skills training were recruited. Formal written informed consent was obtained from each of the students involved in the study. Information was provided about the confidentiality...
of the data and voluntary nature of participation in the study. The study was approved by the AIMST University Ethics Committee and the student’s council as part of the prerequisite for any research involving undergraduate students. To minimise the transference of learning from one manikin to another, the students were initially randomly allocated each of the five manikins, that is, the n=56 were divided into five groups A–E with around 11 students per group. After everyone in the group had intubated within the stipulated 2 min each, the entire group moved to the next station. The groups moved to all five stations to complete the intubations. Patient simulators and airway task trainers that were available at AIMST University were used in the study. Of the manikins used, three, namely the Human Patient Simulator (HPS, V2006), Emergency Care Simulator (ECS V2008) and iStan (V2008), were from Medical Education Technology Inc (METI, USA), which is now absorbed into CAE Healthcare. Two others, Mega Code Kelly (MCK, V2006) and AWTR (V2008), were from Laerdal Medical Corporation (Stavanger, Norway). HPS, ECS and iStan are technologically advanced manikins that can perform physiologically linked full-scale simulations, while MCK and AWTR are part-task trainers. The advanced version of the Laerdal simulators, SimMan 3G, was not available with us at the time of the study.

In each of the five manikin stations, every student was allowed a maximum of 2 minutes to achieve intubation through a conventional Macintosh direct laryngoscopy. Trained faculty stationed at each of these manikins confirmed the ‘definitive intubation’ through auscultation of air entry. These trained facilitators assessed the performance of intubation and recorded the following parameters:

1. ability to achieve successful intubation;
2. evidence of any signs of dental trauma;
3. number of attempts taken for a successful intubation within the 2 min;
4. Cormack Lehane’s grade of visualisation of vocal cords 

reported by the students.

Unsuccessful intubation was declared if a student did not achieve definitive intubation (as confirmed by the faculty) within the stipulated 2 min. Oesophageal intubations were recorded as unsuccessful attempts. ‘Teeth injury’ was documented as ‘present’ if there were ‘click’ sounds on the teeth during laryngoscopy for MCK and AWTR; dislocation of teeth for iStan and splaying of the teeth with manipulation for HPS and ECS. The intubation and laryngeal view during laryngoscopy were recorded as ‘difficult’ if the students expressed a Cormack Lehane laryngeal grade of 3 or 4 and were graded as ‘easy’ if the students expressed grade 1 or 2.

No intubation pillows (to enhance neck flexion), external laryngeal manoeuvres (ELP) or additional airway adjuncts other than stylets were provided for intubation, as these were not in the scope of the undergraduate curriculum. The students were taught the importance of alignment of airway axes, but this was not assessed in the study.

To evaluate the presence of variability in airway characteristics, eight experts, who had more than 5 years of experience in clinical intubations and teaching of airway management skills for undergraduate students, were invited to formally assess (validate) each of the manikins using a 10-point scale (see online supplementary appendix). The data obtained were used to associate with the students’ intubation outcomes. No formal statistical correlations were performed between the expert and student groups as the parameters assessed and experience levels are not comparable.

**Statistical analysis**

Data obtained from student intubation were collated, tabulated and analysed. In our study, each patient had two outcome measurements (success/no success in intubation, presence or absence of tooth injury) from two different aspects of assessment of airways (one was success of intubation and other was teeth injury) and the primary outcome was binary. This was a prospective study and as the outcome measurement was binary, general estimating equation (GEE) analysis for a crossover design was used to compare the different types of manikins. In order to estimate the relative risk (RR), the Poisson model with log-linear type was chosen. GEE analysis was performed using SAS V9.2. After obtaining analysis of GEE parameter estimates and SE estimates, a contrast estimate was performed to make a comparison among the five manikins (RR and 95% CIs).

**RESULTS**

The results obtained (56 students×5 manikins=280 intubations) from the parameters recorded are given below (table 1).

**Successful intubation**

The overall failure rate for all manikin intubations was 25%. The highest rates of inability to intubate were seen in iStan (64.3%) followed by ECS (30.4%), MCK and AWTR (12.5% each). The lowest failure rates were seen with HPS (5.4%).

<table>
<thead>
<tr>
<th>Feature/Manikin</th>
<th>HPS n (%)</th>
<th>ECS n (%)</th>
<th>MCK n (%)</th>
<th>AWTR n (%)</th>
<th>iStan n (%)</th>
<th>Overall n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Successful Intubation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3 (5.4)</td>
<td>17 (30.4)</td>
<td>7 (12.5)</td>
<td>7 (12.5)</td>
<td>36 (64.3)</td>
<td>70 (25)</td>
</tr>
<tr>
<td>Yes</td>
<td>53 (94.6)</td>
<td>39 (69.6)</td>
<td>49 (87.5)</td>
<td>49 (87.5)</td>
<td>20 (35.7)</td>
<td>210 (75)</td>
</tr>
<tr>
<td>Difficult Grade of Larynx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0 (0)</td>
<td>25 (44.6)</td>
<td>8 (14.3)</td>
<td>0 (0)</td>
<td>2 (3.6)</td>
<td>35 (12.5)</td>
</tr>
<tr>
<td>No</td>
<td>56 (100)</td>
<td>31 (55.4)</td>
<td>48 (85.7)</td>
<td>56 (100)</td>
<td>54 (96.4)</td>
<td>245 (87.5)</td>
</tr>
<tr>
<td>Teeth Injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>32 (57.1)</td>
<td>24 (42.9)</td>
<td>0 (0)</td>
<td>8 (14.3)</td>
<td>35 (62.5)</td>
<td>99 (35.4)</td>
</tr>
<tr>
<td>No</td>
<td>24 (42.9)</td>
<td>32 (57.1)</td>
<td>56 (100)</td>
<td>48 (85.7)</td>
<td>21 (37.5)</td>
<td>181 (64.6)</td>
</tr>
</tbody>
</table>

Data showing presence of variability in airway intubation on five manikins. Compared to iStan, there were statistically significant lower difficulty rates with ECS (p<0.0001) and MCK (p<0.0313). iStan had the highest failure to intubate rates compared to other manikins (p<0.0001). HPS and iStan had significantly higher teeth injury compared to ECS, MCK and AWTR (p<0.0001); AWTR, Airway Trainer and iStan; ECS, Emergency Care Simulator; HPS, Human Patient Simulator; MCK, Mega Code Kelly.
The difference in failure rates for intubation between iStan and the rest of the manikins was statistically significant (p<0.0005); RR for HPS-iStan: 0.08 (95% CI 0.03 to 0.24); ECS-iStan: 0.48 (95% CI 0.31 to 0.72); MCK-iStan: 0.19 (95% CI 0.10 to 0.38) and AWTR-iStan: 0.20 (95% CI 0.10 to 0.42).

Tooth injury
Injury to teeth was recorded in one-third of all the intubations (35.4%). iStan intubations were associated with the highest (62.5%) tooth injury rates followed by HPS (57.1%), ECS (42.9%) and AWTR (14.3%). MCK was not associated with any tooth injury. The rates of tooth injury in the iStan and HPS manikins were significantly higher (p<0.0001) compared to those in the other three manikins. RR contrast estimates showed 1.14 for HPS-iStan (95% CI 0.76 to 1.73), 1.52 for ECS-iStan (95% CI 1.06 to 2.18), 2.68 for MCK-iStan (95% CI 1.92 to 3.73) and 2.31 for AWTR-iStan (95% CI 1.63 to 3.28).

Number of attempts taken to intubate
The first-attempt intubation success was achieved in 92.6% of HPS manikins, 87.5% of MCKs, 73% of AWTRs and 46.4% of ECS manikins. Data for the number of attempts on iStan manikins were not recorded by the facilitator in that station and hence statistical analysis or cross tabulation could not be performed.

Grading of difficulty in visualisation of vocal cords
Grades of visualisation of vocal cords were recorded from all 56 participants during intubation. The grade was deemed ‘difficult’ if it was stated by the student as Cormack Lehane grade 3 or 4. The ‘no difficulty’ rate of laryngeal view (Cormack Lehane grades 1 and 2) was 100% with HPS and AWTR manikins, while ‘difficult’ grades were noted with ECS (44.6%), MCK (14.3%) and iStan (3.6%). RR estimates were 1.04 for HPS-iStan (95% CI 0.97 to 1.11), 0.56 for ECS-iStan (95% CI 0.45 to 0.71), 0.88 for MCK-iStan (95% CI 0.78 to 0.99) and 1.03 for AWTR-iStan (95% CI 0.96 to 1.08).

An analysis of GEE parameter estimates and empirical SE estimates showed that in comparison to iStan there were significant ‘no difficulty’ rates with ECS (p<0.0001) and MCK manikins (p=0.0313). However, there was no significant difference for the other two types of manikins when compared to the iStan.

DISCUSSION
Simulation helps to teach core academic content in a hands-on manner, akin to traditional bedside teaching.11 12 Simulators have evolved into a tool to assess individuals and teams in achieving critical objectives required for crisis management.12 The study was designed to explore the presence of variability in airway characteristics in the simulation manikins available with us. Randomising into five groups and spacing the time between intubation on each of the manikins (∼22 min between each manikin intubation by each candidate) was planned to reduce the influence of intubator’s fatigue had they been made to perform intubation on all five manikins in a row. Adequate allocation into groups and spacing the intubation attempts were also aimed at minimising transfer of learning from one manikin to another. Silby et al evaluated four airway task trainers through placement of laryngeal mask airway (LMA) and assessed various parameters. AWTR that was also used in our study was rated to have the best score from participants. The other three manikins used could not be tried in our study as they had used sliced models and were not very anatomically comparable. Our study was on intubating conditions and, in particular, we were keen to evaluate the high-end simulators that are used in more complex scenarios.5

Based on our study results and the airway assessment by the experts, the iStan manikin was the toughest to intubate with the lowest success rates for intubation (35.7%) and highest rates of teeth injury (62.5%). This could be attributed to iStan’s rigid neck with limited neck extension (15°), limited mouth opening (one fingerbreadth) and large anteroposterior diameter of head (>9 cm) resulting in difficulty in intubation. Further analysis showed that the students graded ‘difficulty in laryngeal view—grades 3 and 4’ only in 3.6% of iStan intubations, but had the highest rates of teeth injury and were the least successful with intubations on it. In addition, the upper airway of iStan is not designed to achieve an airtight seal for mask ventilation scenarios.

The study showed that the ECS manikins were graded second in the level of difficult manikins with ‘no success of intubation’ rates of 30% and had the highest ‘difficult grade of visualisation of larynx’ of 44.6%. This was statistically significant when compared to the other manikins. The HPS manikins, like the AWTR manikins, were graded to have the ‘least difficult’ laryngeal view by the students (100% no difficulty), while the HPS manikins had the ‘most successful’ intubation attempts (5.4% failure rates), followed by the AWTR (12.5% failure rates). Subgroup analysis and cross tabulations showed that though the HPS had a favourable intubation profile, this was with the presence of very-high teeth injury (57.1%). A retrospective study on dental injuries after anaesthesia and intubation showed that there was a higher association between difficult intubation predictability and occurrence of dental damage.13

One of the possible causes for airway intubation difficulty in our manikin-based study could be the absence of intubating pillows to achieve neck flexion and the use of external laryngeal manoeuvres like BURP (backward rightward upward pressure) and OELM (optimal external laryngeal manoeuvre). However, as part of the undergraduate curriculum, the students were not taught these techniques during the training periods at the simulation centre. The head extension was kept optimally appropriate by the candidate during intubation. However, this could be very variable in each of the five manikins owing to their relative plasticity and inbuilt components in the neck (e.g., iStan), which can limit full extension.

Lye et al14 showed that novice and trained personnel had no difference in intubation success rates in a difficult airway scenario with simulated airway difficulties on manikins. No recent study has shown the comparison of inherent variability in the airway characteristics of manikins. This is particularly relevant where large numbers of undergraduate students (>150) are rotated on manikins in a row. Adequate allocation into groups and spacing the intubation attempts were also aimed at minimising transfer of learning from one manikin to another. Silby et al evaluated four airway task trainers through placement of laryngeal mask airway (LMA) and assessed various parameters. AWTR that was also used in our study was rated to have the best score from participants. The other three manikins used could not be tried in our study as they had used sliced models and were not very anatomically comparable. Our study was on intubating conditions and, in particular, we were keen to evaluate the high-end simulators that are used in more complex scenarios.5

Based on our study results and the airway assessment by the experts, the iStan manikin was the toughest to intubate with the lowest success rates for intubation (35.7%) and highest rates of teeth injury (62.5%). This could be attributed to iStan’s rigid neck with limited neck extension (15°), limited mouth opening (one fingerbreadth) and large anteroposterior diameter of head (>9 cm) resulting in difficulty in intubation. Further analysis showed that the students graded ‘difficulty in laryngeal view—grades 3 and 4’ only in 3.6% of iStan intubations, but had the highest rates of teeth injury and were the least successful with intubations on it. In addition, the upper airway of iStan is not designed to achieve an airtight seal for mask ventilation scenarios.

The study showed that the ECS manikins were graded second in the level of difficult manikins with ‘no success of intubation’ rates of 30% and had the highest ‘difficult grade of visualisation of larynx’ of 44.6%. This was statistically significant when compared to the other manikins. The HPS manikins, like the AWTR manikins, were graded to have the ‘least difficult’ laryngeal view by the students (100% no difficulty), while the HPS manikins had the ‘most successful’ intubation attempts (5.4% failure rates), followed by the AWTR (12.5% failure rates). Subgroup analysis and cross tabulations showed that though the HPS had a favourable intubation profile, this was with the presence of very-high teeth injury (57.1%). A retrospective study on dental injuries after anaesthesia and intubation showed that there was a higher association between difficult intubation predictability and occurrence of dental damage.13

One of the possible causes for airway intubation difficulty in our manikin-based study could be the absence of intubating pillows to achieve neck flexion and the use of external laryngeal manoeuvres like BURP (backward rightward upward pressure) and OELM (optimal external laryngeal manoeuvre). However, as part of the undergraduate curriculum, the students were not taught these techniques during the training periods at the simulation centre. The head extension was kept optimally appropriate by the candidate during intubation. However, this could be very variable in each of the five manikins owing to their relative plasticity and inbuilt components in the neck (e.g., iStan), which can limit full extension.

Lye et al14 showed that novice and trained personnel had no difference in intubation success rates in a difficult airway scenario with simulated airway difficulties on manikins. No recent study has shown the comparison of inherent variability in the airway characteristics of manikins. This is particularly relevant where large numbers of undergraduate students (>150) are rotated through simulation centres requiring many types of advanced manikins to be used simultaneously to deliver the same academic content. At the time of the study, we did not have the advanced airway trainers with inbuilt learning analytics and video feedback such as the MW11 difficult airway management simulator (Kyoto Kagaku Co, Japan) for comparison. Neither did we have the advanced version, SimMan3G (Laerdal, Stavanger, Norway), for a realistic comparison with HPS, ECS and iStan.

From the student data and expert comments on various manikin airway components, the differences between the manikin airway characteristics can be attributed to significantly limited neck mobility (15°) in iStan with a rigid neck. The mouth opening of iStan is inherently limited (<2 finger’s breadth) by design. The epiglottis is also short and stubby with restricted manoeuvrability in the iStan and ECS manikins.
Intubation might be difficult in this case, especially if only stylets are used as adjuncts compared to bougies to assist intubations. Airway axes alignment with no neck flexion by addition of intubating pillows could have led to varying intubation outcomes. Particularly with the iStan, the rigid neck with minimal head extension might explain the poor success at intubation and high teeth injury in spite of having a very good laryngeal view. The experts’ consensus was that the airways of HPS and AWTR were the most comparable simulators to human anatomy in clinical settings.

The presence of the facilitator directly recording the intubation attempt might have influenced the difficulty of intubation, especially when tooth injury is closely monitored. A more meticulous study design involving remotely monitored teeth injury and servo controlled recognition of definitive intubation with no ‘spectator or peer pressure’ inside the simulation room might help in the ease of intubation. One of the limitations of the study was the potential bias on laryngeal grading stated by the students, influenced by opinions of other students in their group or by watching other students in their group struggling to succeed in intubation. Future studies could possibly address this in two ways: first, by doing an intragroup randomisation or by performing intubation within the confines of a quiet room with only the student and the observing faculty present inside.

Our work provides all simulation educators a basic knowledge of the airway features of these manikins. In one of the author’s institute, where more than 200 undergraduate students performed airway intubations once every 6 months on high-end manikins, there was a neck fracture in the manikin that resulted in a 6-month downtime for replacement and resumption of the training. Most present-day technologically advanced manikins are engineered with capabilities to increase simulated airway training. Most present-day technologically advanced manikins are engineered with capabilities to increase simulated airway tasks difficulty levels by tongue swelling, pharyngeal occlusion, vocal cord partial and complete adduction and trismus. If the manikins have all these programmable specifications, is there a need to have such variability on the inherent airway characteristics as shown by our study? ‘Capturing clinical variations’ within the simulated environments is arguably a cornerstone for simulation-based health education. However, non-programmable and non-quantifiable anatomical variability among different simulator models can affect learning outcomes for students unfairly, especially in the situation of assessments.

The use of simulation has revolutionised summative assessments such as undergraduate final professional examinations, postgraduate examinations and maintenance of certifications in anaesthesia through addition of performance stations to assess clinical competency and soft skills. A standard setting for these ‘active’ stations mandates the application of basic principles of assessment such as validity, reliability and consistency with minimal inter-rater variability, in this case the simulators used. Lampotang described repeatability as a: ‘high level of confidence that any aberrant simulator output/outcome is solely due to the actions or inactions of the clinicians’ and described the concepts of interstimulator and intrasimulator repeatability that can affect the quality of high-stakes decisions in summative assessments. Our study showcases problems with intersimulator repeatability and the existence of inherent problems in standardisation of student assessment when various models of simulators are used to assess airway management skills in high volume coordinated tests like objective structured clinical examinations (OSCE).

Our study will help to make educators to be aware of issues with using different simulator models in high-stakes assessments. Most acute care specialties that introduce crisis management simulations would require airway stabilisation at the very minimum, before they could progress on to the conundrum of differentials and management protocols. It is known that participants who had prior exposure to similar manikin anatomy and design performed better than those not exposed to the same manikin type for the same clinical skill assessed. This can potentially over-rate a few while unduly under-rating the performers who do not have the prior exposure advantage to the manikin model that was used during summative assessments. When assessments are carried out concurrently for a large number of candidates, educational institutes are compelled to use different manikin platforms that have comparable technical specifications to run the scenarios. If existing manikins have such contrasting inherent difficulties in airway characteristics, it could be a cause for concern when we aspire to achieve standardisations in assessments.

CONCLUSION
Heterogeneity in airway features of the manikins used for training students and their assessment prompts the need for standardisation. As a prerequisite, students must have prior component airway training on part-task trainers before exposure to expensive high-technology simulator-based training. The cost of downtime with manikins with an inherent difficult grade of intubations and equipment damage is high.

Uniformity in progression of automated scenarios is influenced by definitive intubations. Educators/trainers who use the manikins for crisis management skills must be aware of these varying features; if not, the quality of learning outcomes from the automated scenarios will deteriorate.

There are concerns in validity of assessment if manikins with inherent variability in airway characteristics are used for summative examinations, physician reaccreditations and OSCE, hinting at the lack of standard setting and inter-rater (in this case manikin) variability.

Acknowledgements The authors would like to acknowledge the AIMST University simulation centre staff for the technical assistance in the smooth conduct of the study and the students for their kind participation. They also extend a special thanks to Associate Professor Eugene Liu and Professor Lee Tat Liang from National University Singapore for their valuable feedback and guidance during the preparation of the manuscript.

Contributors BA, KN, AB, DP, SS and ETL participated in the design, conduct, data collection and preparation of the manuscript. In addition, SL provided methodological support and data refining and statistical analysis. FGC helped conceive the idea of the study and monitored the conduct of the study and revised the manuscript direction.

Competing interests At the time of preparation of the manuscript, one of the authors, AP, was a provider of educational support for CAE Healthcare as adjunct faculty. The rest of the authors declare no conflict of interest.

Ethics approval AIMST University Ethics committee and student advisory council for educational studies performed on undergraduate medical students.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Additional data are attached as an online supplementary appendix.

REFERENCES


