Standards for simulation in anaesthesia: creating confidence in the tools

D. Cumin1, J. M. Weller2,3, K. Henderson3,4 and A. F. Merry1,3*

1 Department of Anaesthesiology and 2 Centre of Medical and Health Sciences Education, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand
3 Auckland City Hospital, Auckland, New Zealand
4 Advanced Clinical Skills Centre, Auckland, New Zealand
* Corresponding author: A. F. Merry. E-mail: a.merry@auckland.ac.nz

Key points
- Standards are important to ensure confidence in results of simulation-based education, research and assessment.
- There are few standards related to simulation.
- Standards should be objective and testable.
- Broad consensus is needed to develop relevant and useful standards for simulation.

Summary. Simulation is an accepted part of training, assessment, and research in aviation, nuclear power, and the military. Confidence in results in these industries is underpinned by relatively comprehensive and widely accepted standards. In contrast, although there have been major advances in the technology and tools used for simulation in the healthcare industry over the last few decades, little work has been done in setting standards for simulation in healthcare. Standards are essential for achieving the full potential of simulation-based education, assessment, and research at all levels and specialities in healthcare. The absence of standards undermines confidence in the results of any simulation-based endeavour and increases the risk of negative learning. We propose a practical framework for setting standards for simulators for anaesthesia.

Keywords: equipment, simulators; equipment, standardization; model, computer simulation; physiology

The advantages of simulation for training, assessment of competence, and research and development have become increasingly clear in industries such as aviation, nuclear power, and the military (Table 1).1–3 It is widely believed that simulation may offer similar benefits to healthcare.4–8 Despite debate about just how beneficial simulation is9–12 and the difficulties in evaluating its effectiveness,13 simulation in medicine (particularly anaesthesia) is gaining momentum and is now a compulsory part of training in some countries.14–15 In the UK, the Chief Medical Officer has called for an increase in the use of simulation for training medical personnel.16 In Australia and New Zealand, The Australian and New Zealand College of Anaesthetists (ANZCA) established the Effective Management of Anaesthetic Crises (EMAC) course in 2002,17 which is now a training requirement and widely used in continuous professional development. There is interest internationally in extending the use of simulation to the assessment of doctors, and this has already begun in some centres.18–19 Simulation is also increasingly used in certain types of research, often with the objective of improving patient safety.20–22

In each of these applications, there is reliance on the ability of actors or devices, typically within a complex simulated environment, to create an authentic experience replicating a live clinical encounter. High fidelity in simulation does not in itself assure effective education, valid and useful research outputs, or reliable assessment. Conversely, it is of course possible to provide effective education, undertake worthwhile research, and reliably assess certain aspects of clinical competency using basic simulators, actors, or standardized patients—or indeed without simulation at all. Nevertheless, for certain objectives, the authenticity of the simulated experience may be very important. For example, much simulation-based anaesthesia research is predicated on the notion that behaviour in a simulated scenario will reflect behaviour in clinical settings, and generally the plausibility of this assumption increases with the realism of the experience. Similar comments apply to educational exercises in which the objectives include the promotion of effective performance in complex clinical settings, and to high-stakes assessment.

Authenticity depends on many elements of simulation, including the staff, the setting, and the simulator itself. Furthermore, pedagogical and methodological considerations are of critical importance in achieving worthwhile goals.23–26 Those responsible for simulation centres can control most of these elements, but they are dependent on manufacturers for the quality of the simulators they use. When purchasing a simulator, they will expect certain features (determined in accordance with their particular objectives) simulated to a reasonable level of fidelity. Unfortunately, they may be disappointed: in currently available simulators, physiological models are often unrealistic, and even simple physical features (such as radial pulses) may be deficient.27–29 This may create challenges to authenticity that can be difficult to manage, and thereby undermine attention to detail in the other aspects of simulation.

There is a good argument for setting standards in simulation.25–28 The Society for Simulation in Healthcare (SSH)
has recently published SSH Accreditation Standards for Simulation Programs in Healthcare. These standards outline criteria for key aspects of simulation-based teaching, assessment, and research and provide additional accreditation for the promotion of patient’s safety. While these standards are a start, they are rather general in nature. For example, one criterion for educational accreditation of a learning centre requires that ‘simulation modalities are appropriate for the learning objectives’, but there is no specific indication of what constitutes appropriateness. Standards should define the boundaries of what is required, expected, desired, or acceptable and thereby provide confidence in products, personnel, or processes. In the absence of specific detail, deciding whether a standard has been achieved becomes subjective, and a matter of opinion.

The comparison with aviation and nuclear power is marked. These industries have standards that promote reproducibility (i.e. one simulator should be very much like another of the same type and should consistently produce the same responses to given inputs), and define benchmarks to provide some surety to users that a simulator will function to the required standard. Standards in simulation for aviation are regularly updated—a process that serves to constantly improve simulators and their use, taking into account new learnings from events such as airline crashes and power plant disasters. An example of this was provided by the 2001 American Airlines flight 587 crash which killed 260 passengers and five people on the ground. The investigation into the accident found that pilot error was, at least in part, to blame for the accident. The pilot had been trained in a simulator to use a manoeuvre that contributed to the crash. The investigation found that limitations of the simulator acted to reinforce negative learning. Among other things, it was recommended that training programmes note the limitations of simulators and restrict teaching of important manoeuvres in simulators unable to adequately reproduce those manoeuvres to an ‘academic briefing’ only.

In a similar way, regularly updated standards are needed for all aspects of simulation in healthcare, including the curriculum, the staff, the environment, and the methods of teaching, research, or assessment. Some standards should apply widely (e.g. those related to teaching techniques), but others will be specific to disciplines: simulation is well established in psychiatry, for example, but does not depend on the sophisticated models of the pharmacology of drugs relevant to many scenarios in anaesthesia: it follows that some standards for the latter will be of little relevance to the former.

Developing comprehensive standards will take many years. In this paper, we focus on one aspect of this task: standards for simulators in anaesthesia. We build on a previous systematic analysis of available simulators related to anaesthesia and propose an approach which could perhaps be generalized to other aspects of simulation and to other medical disciplines.

### Standards for simulators

Several classifications have been proposed for simulators. These typically focus on levels of fidelity or the ways in which users interact with the simulator. We have previously reviewed these classifications and proposed an alternative approach, based on three key attributes of anaesthesia-related simulators: their use, method of interaction, and physiology. This approach lends itself to standard setting.

A key challenge in developing standards for anaesthesia simulators lies in the potentially overwhelming nature of the task: to specify in detail the requirements for every feature on every different type of simulator would be virtually impossible. However, many of the features found on the most complex simulators are also represented in isolation or in combination with a smaller set of other features on simpler simulators. A simulated mouth and larynx, for example, may be found on a simple airway trainer or in combination with simulated ECG traces and other features on a complex simulator intended for immersion training. The simulated ECG traces may similarly be found in isolation on a screen-based ECG recognition trainer. While objectives of any simulation should guide decisions as to which features are required, the standards for each required feature would be the same whether the feature was presented in the form of a simple trainer or as part of a more comprehensive simulator. The requirements to teach or assess the task of tracheal intubation, for example, are the same whether this is being done in isolation or as a part of an immersion scenario in which teamwork and communication are also being assessed. A refinement of this position would recognize that the level of fidelity required of a given feature may vary according to the objective of the simulation. This variation occurs on a continuum from absent to highly realistic, but for pragmatic reasons, we propose an adaptation of an approach used by the aviation industry to set fidelity levels for flight simulators. This would involve defining three levels of fidelity: S, specific, realistic, and meeting defined standards; R, representative, simply present, and recognizable; and A, absent. It is not necessary for every possible

### Table 1 Advantages of simulation for training, assessment of clinical skills, and certain types of research. Adapted from refs 1 2 37 44 48 63

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation reduces the risk and inconvenience to participants, the public inherent, or both in conventional research and training</td>
<td>Simulation allows training exercises to be standardized, repeated, and tailored to the needs of individuals</td>
</tr>
<tr>
<td>Simulation allows the environment to be manipulated and scenarios to be controlled, so that responses to specified (often uncommon) events can be practised</td>
<td>Simulation can usually be stopped, sped up, or slowed down to facilitate the achievement of certain learning objectives</td>
</tr>
<tr>
<td>Observation and recording of ‘clinical’ events and human performance, with the opportunity for feedback and reflection in a safe environment, is facilitated by simulation</td>
<td>Suitable scenarios to be controlled, so that responses to specified (often uncommon) events can be practised</td>
</tr>
</tbody>
</table>

* BJ A* Cumin et al.
Table 2: Illustrative standards for some interactive features of anaesthesia simulators:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Standard</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Chest</td>
<td>1.R The chest should be similar in size and shape to that of a healthy adult</td>
<td>The chest should be easily recognized as such</td>
</tr>
<tr>
<td></td>
<td>1.S.1 As above</td>
<td>As above</td>
</tr>
<tr>
<td></td>
<td>1.S.2 The chest should simulate breathing</td>
<td>The chest should rise and fall at the ventilatory frequency and breath sounds be present (see standard for breath sounds below)</td>
</tr>
<tr>
<td></td>
<td>1.S.3 It should be possible to use a real defibrillator on the chest</td>
<td>Defibrillator pads should transfer shocks from the defibrillator to a voltage measuring device in the chest</td>
</tr>
<tr>
<td></td>
<td>1.S.4 ECG leads should be attachable to the manikin via electrodes in the positions in which they are placed on a real patient</td>
<td>Leads that resemble standard ECG leads should be available and be readily attached to electrodes on the chest or the chest itself. Various waveforms should be recognizable (see standards for ECG display)</td>
</tr>
<tr>
<td></td>
<td>1.S.5 Cardiopulmonary compressions should be possible as they would be on a real patient</td>
<td>The chest should be depressible to at least one-third of its antero-posterior diameter with a constant resistance</td>
</tr>
<tr>
<td>2. Breath Sounds</td>
<td>2.R Bilateral breath sounds that can be auscultated should be present</td>
<td>Bilateral breath sounds should be recognizable as such. Breath sounds should be synchronized with the rising and falling of the chest</td>
</tr>
<tr>
<td></td>
<td>2.S.1 Unilateral breath sounds that can be auscultated should be present</td>
<td>Unilateral breath sounds should be recognizable as such</td>
</tr>
<tr>
<td></td>
<td>2.S.2 A range of breath sounds should be present (e.g., normal, crepitations, rhonchi).</td>
<td>A specified list of breath sounds should match a set of exemplars, tested by dynamic time warping (an algorithm for measuring similarity between two time-series sequences that is particularly relevant to sound comparisons). Switching between the sounds (e.g., from normal to crepitations or rhonchi) should be easily controlled</td>
</tr>
<tr>
<td></td>
<td>2.S.3 The volume should be controllable by the instructor</td>
<td>Volume should be controllable in 5% increments between 0 and 1 dB</td>
</tr>
<tr>
<td>3. Display of three-lead ECG</td>
<td>3.R.1 Sinus rhythm should be available for display</td>
<td>It should be possible to display a waveform that matches a sample sinus rhythm (tested with an appropriate time-series technique). The distance between peaks of the QRS complex should be equal to the heart rate frequency (see standards for heart rate) and synchronized with the palpable pulse. The data on the monitor should be refreshed at a rate of at least 30 fps.</td>
</tr>
<tr>
<td></td>
<td>3.S.1 As above</td>
<td>As above</td>
</tr>
<tr>
<td></td>
<td>3.S.2 The following rhythms should be present: sinus, sinus bradycardia, sinus tachycardia, ventricular fibrillation, nodal rhythm ventricular tachycardia, and asystole</td>
<td>The ECG should match a specified list of sample waveforms using a specified time-series technique, and should be within 5% of the specified rate</td>
</tr>
<tr>
<td>4. Central venous catheter</td>
<td>4.R It should be possible to place a central venous catheter in the neck as if into an internal jugular vein</td>
<td>There should be a hole in the neck of the simulator just large enough to accommodate a real central venous catheter</td>
</tr>
<tr>
<td></td>
<td>4.S As above, with a central venous pressure trace relayed to an external monitor</td>
<td>As above. Pressure from the central venous catheter should match the central venous pressure specified by the simulator to within 5%</td>
</tr>
</tbody>
</table>

Feature to be present for every objective of simulation, and it can be argued that for some objectives, it may even be an advantage for certain features to be missing to limit the complexity of a simulation and thus avoid distracting or overloading a novice participant. For example, an ECG might serve as an unhelpful distraction to a novice learning to intubate the trachea, and in this context, it might also be advantageous to designate the chest wall as ‘absent’ to facilitate visualization of the effects of ventilation of the lungs once the tracheal tube is in the correct place.

Using this approach, standards for simulators could be framed in reference to their interactive features (the physical features with which participants interact and the way in which the simulator interacts with ancillary equipment), and their physiological modelling. Standards need to specify the requirements for such features and should also include tests for each feature that could be used to verify satisfactory compliance (Table 2).

Simulated physiology may be model-based, script-based, or a mixture of these modalities. Script-based simulators are subject to operator error and bias. Model-based simulators theoretically obviate these problems, but the modelled physiological responses of available simulators to certain clinical events and drugs are not all equally satisfactory. The examples (Table 2) are typical of some of the basic requirements for any complex simulator, but the test definitions do not address the more sophisticated aspects of modelled physiology and pharmacology that one might expect from a comprehensive simulator intended for use in immersion scenarios or of a screen-based simulator for...
teaching complex skills, such as the physiological responses of the ECG and central venous pressure to the administration of i.v. and inhaled drugs to induce and maintain anaesthesia. These responses also require standards.

One approach to evaluating simulated physiology and pharmacology is to obtain ratings of realism from participants in simulations.48–50 We have adopted a more formal test of realism using a modification of the Turing test (from Turing’s ‘imitation game’).51 This involves giving clinicians electronically generated physiological records of a simulation and of a real case and asking if they can distinguish the one from the other: if most cannot, the physiological record of the simulation would be deemed adequately realistic (or conversely).

A standard for simulated model-controlled physiology might specify the range of acceptable values for a particular variable during a particular intervention (or set of interventions) for a particular patient (Table 3). One could specify these on the basis of published pharmacokinetic and pharmacodynamic data (e.g. the Marsh52 or Schnider53 model for propofol). However, such data are not always available, accurate, or applicable to a given simulated patient. Furthermore, these models are often only applicable to specific patient groups and may not be easily applied to interventions where multiple drugs interact.

Objectively, testing the physiological responses of simulators to particular interventions is also made more difficult by the generally non-linear nature of the time-series signals and the complex inter-relationships between physiological variables. Techniques for signal comparisons need to be evaluated, and there may be a need for new techniques to overcome these challenges. The aviation industry uses both subjective and objective tests in the qualification of flight simulators and using a combination of methods seems sensible for medical simulators.

The repeatability of a simulator’s response to an intervention (such as the administration of a specific dose of propofol) is also important. A major advantage of simulation in education, research, and assessment of performance is the ability to standardize, schedule, and repeat events (Table 1). It is true that similar patients, or even the same patient on separate occasions, may respond in different ways to similar interventions. However, for most purposes, the response to a specified simulated event should be essentially identical time and time again. If a particular simulator responds variably and unpredictably to a given intervention, confidence in results of a research study using that simulator, or in the validity of its use in education or assessment, will be limited. Data suggesting a lack of reproducibility have been presented for at least one simulator.54 A standard should perhaps reflect the inherent variability of patients and balance realism in this regard with the benefits of reproducible scenarios. As a minimum, it should be possible, using the same simulator, to provide the same experience to multiple participants, within defined limits of variance. If simulation is to be used for assessment, a participant should be assessed on an experience that is the same as (or at least very similar to) that of other participants taking the same test (provided they use comparable clinical approaches).

Armed with such standards, a task analysis could then be undertaken for any planned simulation.55 56 the required features listed, and the appropriate level of fidelity specified. A simulator could be chosen which not only has the required features, but also meets the relevant standards for those features. Manufacturers would be able to clearly show which features of a simulator meet which standards and consumers could be confident that their purchases would perform as expected.

### Setting and implementing the standards

A list of physical and physiological attributes and features reasonably expected of simulators for specific educational or research-related objectives would be a useful starting point. It would then be important to understand which features are (and are not) important for any given simulation.57–59 However, rather than identify all the features required for each objective, it might be useful to group features (and their standards) together into those required for specific tasks. Simulator manufacturers could then claim to reach standards for specified tasks by reaching the standards for each feature required of each subtask. For example, an objective such as that of assessing a trainee anaesthetist to safely anaesthetize and manage a patient who then develops atrial fibrillation (AF) may require the participant to complete a number of tasks such as

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intervention</th>
<th>Specific standard</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart rate</td>
<td>Baseline</td>
<td>A heart rate should be displayed</td>
<td>The base rate must be within 60 and 80 beats min⁻¹</td>
</tr>
<tr>
<td>Arterial pressure</td>
<td>Baseline</td>
<td>Systolic and diastolic arterial pressure should be displayed</td>
<td>The baseline systolic arterial pressure should be within 100 and 130 mm Hg; baseline diastolic arterial pressure must be within 60 and 80 mm Hg</td>
</tr>
<tr>
<td>Arterial pressure</td>
<td>Administration of propofol</td>
<td>Diastolic and systolic arterial pressure should decrease with induction using propofol</td>
<td>The arterial pressures should decrease to within acceptable limits of a specified model</td>
</tr>
<tr>
<td>Heart rate</td>
<td>Administration of propofol</td>
<td>There should be a transient tachycardia</td>
<td>The heart rate should vary within acceptable limits of a specified model</td>
</tr>
</tbody>
</table>
intubation, induction of anaesthesia, maintenance of anaesthesia, and recognition and management of the heart rhythm (Fig. 1). A simple simulator may be sufficient to teach or assess each of these tasks individually. The recognition of AF, for example, could be performed on a simple screen-based simulator meeting the standard for the task of recognizing heart rhythms. However, the stated simulation objective requires both the recognition and management of the AF, a more complex matter. It may be argued that a model-based simulator is required for assessment of performance, in the interests of objectivity. However, even if a script-based simulator is chosen, the standards for the mouth and larynx, ECG display, and chest should be met. In a model-based simulator, we would also expect the physiological response of the simulator to interventions to meet certain standards.

It can be seen how grouping standards around tasks would simplify matters. A simulation centre buying a simulator that meets standards for the management of AF, for example, will know that it can be used for a variety of scenarios also involving the recognition of this arrhythmia. Grouping features in this way will require careful thought, and will depend to some degree on the economics of simulation design and manufacture. Manufacturers may offer alternatives that meet only part of a given standard, but this deficiency would then be explicit. Alternatively, they may offer simulators that exceed the requirements of the standard, and this would be something for them to market and purchasers to decide upon.

The goal should be to identify a manageable number of tasks and tests by which one could be reasonably assured that given simulators were fit for purpose. With increasing sophistication, factors such as the effect of age or weight on the response to certain drugs might be taken into account. An independent body such as an existing simulation society or College of Anaesthetists should, ideally, take responsibility for regularly testing products to ensure compliance.

### Discussion

We have advocated standards for simulation and specifically for simulators. We have provided an overview of a possible approach to setting standards for simulators with illustrative examples. Although the development of standards for even some commonly used simulators would clearly be a major undertaking, we think it is time this task was undertaken on an international basis.

A limitation of our paper is its essentially illustrative emphasis. To actually develop standards along the lines proposed would require the development of consensus between manufacturers and relevant expert user groups, and would be a substantial task. However, a start could be made, some preliminary basic standards developed and promulgated, and then additional standards added over time. The status quo leaves it difficult to evaluate or compare results from studies based on simulation, to make valid inferences regarding individual performances in evaluations, or to be confident that learning in the simulator will apply to real-world situations. Indeed, there is a risk that simulation-based training conducted without standards (or following other stated guidelines) may occasionally have a negative educational impact, as was evident in the aviation example given above.

Some organizations have made a start to setting standards in simulation. ANZCA made a pioneering start with guidelines for running the EMAC course. The set of standards proposed by the SSH is also a large step in the right direction. However, neither the ANZCA nor the SSH standards include specific standards related to providing credentials for the simulators themselves. We should learn from industries such as aviation where ‘the basis for the credibility that the Simulation Industry enjoys today is attributable to the efforts of a relatively small number of airlines to establish common standards for flight simulation’. With continued efforts, the medical simulation industry can similarly build a high level of credibility.
Future directions
To implement, maintain, and enforce standards for simulators, widespread cooperation is needed from the anesthesia colleges and societies, relevant academic institutions, industry, and organizations such as the Society in Europe for Simulation Applied to Medicine (SESAM),62 and the SSH. Standards will go a long way to ensuring that simulation-based research is conducted with rigour; testing of medical devices in a simulated environment can be relied upon to give useful results in the clinical setting; teaching utilizes the potential of simulation to truly enhance healthcare education; and simulation-based assessment of healthcare practitioners and teams is objective, relevant, and reliable. In the end, the goal is improved quality of care for our patients.

Conflict of interest
A.M. has financial interests in Safer Sleep LLC and responsibility for a simulation centre.

Funding
D.C. was supported by the Rolphe and Eve Seelye Scholarship in Anaesthesiology.

References
1 Good ML. Patient simulation for training basic and advanced clinical skills. Med Educ 2003; 37: 14–21
19 Rosenblatt MA, Abrams KJ. The use of a human patient simulator in the evaluation of and development if a remedial prescription for an anesthesiologist with lapsed medical skills. Anesth Analg 2002; 94: 149–53
22 Weller JM, Merry AF, Robinson BJ, Warman GR, Janssen A. The impact of trained assistance on error rates in anaesthesia: a simulation-based randomised controlled trial. Anaesthesia 2009; 64: 126–30
23 Salas E, Burke CS. Simulation for training is effective when . . . Quality Saf Health Care 2002; 11: 119–20
32 Goldsmith TE, Johnson PJ. Assessing and improving evaluation of anesthesiology simulation. Simulat Healthcare 2006; 1: 122–6
33 Airbus. Airbus submission to the national transportation safety board for the American Airlines Flight 587 Belle Harbour, New York, November 12, 2001. 2004
48 Boyd KB, Olivier J, Salameh JR. Surgical residents’ perception of simulation training. Am Surg 2006; 72: 521–4
51 Turing AM. Computing machinery and intelligence. Mind 1950; 59: 433–60
53 Schneider TW, Minto CF, Shafer SL et al. The influence of age on propofol pharmacodynamics. Anesthesiology 1999; 90: 1502–16
60 Grant LJ. Regulations and safety in medical equipment design. Anaesthesia 1998; 53: 1–3