Published in final edited form as:

Simul Healthc. 2015 December ; 10(6): 331–335. doi:10.1097/SIH.0000000000000131.

Unlike history, should a simulator not repeat itself?

Samsun Lampotang, PhD [Professor]

Department of Anesthesiology, University of Florida College of Medicine, Gainesville, FL

The concept of simulator repeatability and a means for quantifying it are introduced by Cumin, Chen, and Merry in this issue.¹ The authors evaluated different methods of objectively evaluating simulator output repeatability and recommended the Median Absolute Performance Error (MDAPE) method based on their analysis; two simulators, a screen-based simulator (Anesthesia Simulator, Anesoft), and a high-integration mannequin patient simulator (Human Patient Simulator, HPS, CAE/METI) were evaluated for repeatability.

What is repeatability?

Repeatability is a measure of the similarity in the outputs of a simulator during repeated runs of a given scenario with identical (high precision required *and*, ideally, high accuracy, as well) inputs, interventions, and events at the exact same times. The given, carefully selected scenario is run with identical initial conditions, patient demographics, and model parameters. Examples of inputs, interventions, and events could be administered drug dose, increase in inspired oxygen concentration, and hemorrhage, respectively. Tight specification, control, and verification of the quality (identical timing, precision, and, where appropriate, accuracy) of inputs to simulators are of paramount importance when conducting a repeatability analysis on the output of simulators.

When is repeatability relevant?

It is important to properly frame the concept of repeatability and to be discerning and nuanced with regard to the context of where repeatability may or may not apply or may even be inappropriate within the domain of simulation in healthcare.

Repeatability is a desired attribute during summative assessment, especially high-stakes assessment where there must be a very high level of confidence that any aberrant simulator output/outcome is solely due to the actions or inactions of the clinician(s) and not to the lack of repeatability of the concerned simulator. Repeatability is relevant for scenario portability (discussed later). Repeatability is also powerful when it can significantly reduce the sample size to adequately power simulator-based studies such as using a simulator to evaluate the efficacy of new features in medical equipment. The sample size was significantly reduced to 17 anesthesiologists for a Human Patient Simulator-based study demonstrating that an

Reprints: Samsun Lampotang, PhD, Department of Anesthesiology, University of Florida College of Medicine, 1600 SW Archer Road, PO Box 100254, Gainesville, FL 32610 (slampotang@anest.ufl.edu).

Financial Disclosure: The author of this editorial is a co-inventor of the Human Patient Simulator (HPS), a simulator analyzed for repeatability in this issue's article by Cumin, Chen & Merry that this editorial accompanies. The author has received a fraction of the HPS royalties that the University of Florida has collected from CAE/METI.

electronic anesthesia machine's Low Flow Wizard (a.k.a. econometer) provided an average reduction of 42.9% in volatile anesthetic consumption rate.² Interestingly, this study was never attempted as a clinical study because multiple confounding influences such as the patient, the surgical stimulus, and the anesthetic time profile and duration would have required a prohibitively large sample size. In this case, the study protocol exploited scenario repeatability to effectively remove these confounding influences.

However, counter-arguments can be made against repeatability for routine education and training. If a scenario unfolds in exactly the same way, trainees could potentially learn to "game the system" by memorizing and rote learning (consciously or unconsciously) a correct response to a given scenario rather than acquiring the learning objective(s) the scenario was intended to convey. Also, the different types of patient variability that are clinically encountered argue for non-repeatable simulators that actually have different responses for identical inputs (discussed later).

Modeling vs. simulation

The terms "modeling" and "simulation" are not the same. The differences can be helpful in developing a nuanced understanding of different types of repeatability. A model, also referred to as a mathematical model, generally consists of mathematical equations that attempt to predict how a system or sub-system will behave. Mathematical models are the engines that power a model-driven simulator and have no physical instantiation, being reduced to mathematical abstractions. On the other hand, a model-driven simulator generally has a physical/visual component that externalizes the mathematical output, ideally in a clinically authentic manner, e.g., actual monitoring equipment used during simulated anesthesia with a high-integration mannequin patient simulator or a graphical depiction of the display of emulated monitoring equipment in a screen-based simulator (Fig. 1).

As a concrete example, equations in a model could predict a time-varying output such as invasive arterial blood pressure (IABP). If the equation is implemented using the formula function in a spreadsheet, the time-varying output is a column of numbers in units of millimeters of mercury (mmHg). However, that model output would not be engaging and would have limited value for hands-on clinical training, as it is entirely numerical. The IABP numerical model output becomes much more useful and realistic if, via a simulator, it is externalized by being continuously displayed in the format that clinicians are familiar with, i.e., as a dynamic, real-time waveform with derived numerical values for heart rate and systolic, diastolic, and mean ABP (Fig. 1). Later, we compare how this externalization of the IABP model output is performed in a screen-based simulator and a high-integration mannequin patient simulator, as these are the two types of simulator examined for repeatability in Cumin et al.

Deterministic models and repeatability

A deterministic model consists solely of deterministic equations that always produce the exact same output if the inputs and interventions (magnitude, direction, timing) are identical. A concrete example of a deterministic model is the equation: heart rate (HR) $\times$ stroke volume (SV) = cardiac output (CO). It is an exact equation; if HR = 100 bpm and SV = 50

ml, then CO will always be exactly 5000 ml/min or 5 l/min and therefore 100% repeatable. The model output CO is always repeatable (exactly the same) between runs even as the inputs change in a dynamic simulation. For example, if SV drops to 25 ml with HR remaining at 100 bpm, every time this new condition is run, the output will always be 2.5 l/min. The simple example above demonstrates the deterministic and repeatability properties that also apply to more complex deterministic models. Deterministic models are, by definition, 100% repeatable and are used in both screen-based and high-integration mannequin patient simulators.

Screen-based vs. high-integration mannequin patient simulators

This author uses “high integration” instead of “high fidelity” to describe mannequin patient simulators like the HPS because fidelity (like beauty being in the eye of the beholder) is subjective to the application and learning objectives for which a simulator is being used, whereas integration is factual. Unlike a screen-based simulator, a high-integration mannequin patient simulator is tightly integrated with actual, unmodified equipment like an anesthesia machine and its associated physiological monitoring equipment. To be externalized to the trainee, the numerical output of the model has to be converted to the format expected by the actual equipment. For example, the IABP data must be transmitted to an IABP module not as numbers but via a digital-to-analog (D/A) converter as a time-varying analog electrical voltage that is properly calibrated (offset and gain in mV/mmHg) for a given make and model in the list of commercially available IABP modules compatible with the simulator.

Similarly, if the model predicts an end-tidal carbon dioxide level (ETCO₂) of 60 mmHg, an exhaled gas mixture containing actual CO₂ that peaks at the end of exhalation at 60 mmHg partial pressure must be physically generated in real time. As a result, the actual gas analyzer can sample the gas mixture and display an ETCO₂ value of 60 mmHg on the associated multi-parameter physiological monitor. Real-world artifacts are possible at (a) the interface between the actual equipment and the simulator such as leaks in breathing circuits and gas sampling lines, electrical noise, and (b) in the actual equipment such as miscalibrated vaporizers, gas flow meters, gas analyzers, blood pressure monitors, and signal drift. Both sets of artifacts are *extrinsic* to the high-integration simulator (Fig. 1). Such real-world artifacts associated with real physical equipment decrease repeatability of the simulated environment and are not present in a screen-based simulator. As an additional degree of freedom, the actual monitoring equipment integrated with the simulator (and extrinsic to the simulator) may be of different models and manufacturers. Therefore, it is important to note and record the serial number and other unique identifiers of equipment and accessories used for evaluating repeatability.

In a screen-based simulator, instead of being physical, the equipment and associated displays are emulated and the patient is represented by a graphical depiction. Model output such as IABP data and the ETCO₂ value are displayed directly (without any need for conversion and transmission) by simply writing to the screen of the emulated monitor an IABP trace (and its derived numerical values) and the number 60, respectively. Similarly, user input to the model of a screen-based simulator such as a minute ventilation (MV) of 7

l/min is directly communicated without concern about the introduction of artifacts. A screen-based simulator is self-contained within its entirely virtual world and is insulated from real-world physical artifacts, faults, and errors (Fig. 1) unless they are being deliberately simulated such as a stuck-open expiratory valve leaflet in an anesthesia circle breathing circuit. Therefore, repeatability should be perfect in a screen-based simulator, as Cumin et al found.

Incidentally, the technique of supplying simulated inputs to actual equipment, such as a physiological monitor, is termed “stimulation.” An alternative to stimulation is a general purpose computer that emulates the graphical display of a physiological monitor, as in some mannequin patient simulator designs. Stimulation is generally more expensive but is 100% realistic as far as the actual equipment is concerned; it allows users to safely learn the “buttonology” to navigate clinical equipment with complex user interfaces such as deep, nested sub-menus. Stimulation also facilitates simulator-based usability studies of prototypes and working models of medical equipment in realistic clinical settings by actual users before they are FDA approved.

Types of repeatability

There are many different types of repeatability. According to our previous differentiation between a model and a simulator, *model repeatability* characterizes the numerical output of the model, not the simulator. Conversely, *simulator repeatability* examines the output of the simulator such as the similarity at a mannequin’s airway in the physical ETCO_2 value (ideally with laboratory equipment and independently of clinical gas analyzers that may be part of the simulated environment) during identical runs. *Simulated environment repeatability* uses the values displayed on the actual clinical equipment integrated with the simulator as the output to be analyzed for repeatability; it includes any errors, artifacts, and/or unintended variations introduced by the actual equipment.

As an example, the internal gas exchange module of an HPS may be consistently generating an ETCO_2 value of exactly 60 mmHg during each identical run. Assume there is an undetected leak (that varies in magnitude depending on how the endotracheal tube is attached relative to an elbow connector housing the gas sampling port) at the gas sampling line of the actual gas analyzer. Variable amounts of air will be aspirated during each run, creating variable room air dilution and thus poor repeatability of the falsely low ETCO_2 value. This example illustrates how simulated environment repeatability may be low while model and simulator repeatability themselves are both high because an artifact has been introduced into the simulator output.

Simulator input can also contain real-world artifacts that affect both simulator and model repeatability. For example, assume the ETCO_2 value is supposed to be 40 mmHg at an MV of 7 l/min as an input. With an undetected leak of varying magnitude in the breathing circuit or from a variably underinflated endotracheal tube cuff, the lower than specified MV (hypoventilation) will lead to non-repeatable ETCO_2 values greater than 40 mmHg. The input error (lower than specified MV) will also affect the physical gas composition in the simulated alveolar space. The gas analyzer internal to the HPS samples the alveolar space,

providing an artificially high CO₂ value (as a result of the unintended hypoventilation) as a model input, which in turn results in deviation from the expected model output. The circuit leak example highlights the need to verify the integrity, accuracy, and precision of the actual clinical equipment, such as performing an anesthesia machine pre-use check, before performing *each* run of a repeatability analysis. In a screen-based simulator, model, simulator, and simulated environment repeatability should, in principle, all three be the same because there is no need for output conversion, transmission, and actual equipment and the artifacts and errors they possibly introduce.

Intra-simulator repeatability compares the output *within* a given simulator (one actual physical piece of equipment or one specific software package) for identical inputs while inter-simulator repeatability addresses outputs *between* different simulators.³ Inter-simulator repeatability itself can be of many types and is listed in increasing degrees of separation: (a) *intra-design* - identical design (identical software version, model, and manufacturer); (b) *intra-type* - same type (e.g., mannequin patient simulator) but of different models and manufacturers, and (c) *inter-platform* - different platforms (e.g., screen-based platform vs. high-integration mannequin simulator platform).

It is therefore good practice to always specify which repeatability is being discussed. For evaluating intra-simulator repeatability of a high-integration simulator, inputs must be high precision while accuracy should ideally also be high but not necessarily so. For evaluating inter-simulator repeatability of high-integration simulators connected to different input devices such as anesthesia machines of different models, manufacturers, and designs (such as bellows vs. piston ventilator), the inputs must be both high precision and high accuracy. Low accuracy between and within input devices leads to inputs to the simulator that are actually of different values instead of identical, artifacts that make repeatability appear lower than it actually is.

Intra-simulator repeatability is relevant for situations like scenarios used for high-stakes assessment. If more than one simulator of the same model is used for high-stakes exams, then intra-design, inter-simulator repeatability is also important. Inter-simulator repeatability is also relevant for scenario portability; that is, the ability to share fully validated scenarios among institutions nationally and internationally rather than re-invent the scenarios with the caveat that any race-specific modeling must be addressed for international portability. Portability through inter-simulator, intra-design repeatability between different simulators of the same commercial model is already attainable.⁴ However, inter-simulator, intra-type repeatability and thus portability between commercial simulators from different manufacturers connected to input devices of different model, manufacturer, and design will require more effort. It is also important to note that depending on what model parameters are crucial for a given scenario, a given scenario A may be repeatable and thus portable between simulators X and Y but another scenario B may not be repeatable and thus not portable between simulators X and Y. Therefore, the choice of scenarios used to assess repeatability is important.

Cumin et al. analyzed intra-simulator *model* repeatability of the HPS. The HPS model output, instead of the simulated environment output displayed by the actual monitors, was

used to assess repeatability, adopting a model-centric, instead of a clinician- or user-centric, approach. For a high-stakes exam, simulated environment repeatability is more relevant than model repeatability. The former is the simulated reality that the clinician being assessed sees and experiences, not model repeatability, which is usually not made available or displayed to clinicians/trainees.

The examples above illustrate why repeatability in a high-integration simulator stimulating actual monitoring equipment and receiving input from actual medical devices cannot be expected to be at the level of a screen-based simulator, as Cumin et al. also found. Variability (deliberate or inadvertent) in the input to a simulator will affect repeatability of the simulator output. This is true for both model- and instructor-driven simulators. The potentially larger input variability between instructors (e.g., varying levels of up-to-date knowledge and understanding of physiology and pharmacology and the ability to think on their feet) could result in lower repeatability in instructor-driven simulators compared to model-driven ones.

Variability vs. repeatability

There are many parallels between simulation in healthcare and flight simulation. One major difference is that what we do not know or cannot measure about a single patient, let alone the entire world population, exceeds what we do know and can measure; this will likely remain true for the foreseeable future, if not forever. Also, some of our knowledge of the human body is in the form of associations instead of quantitative cause and effect relationships; the latter are more useful and readily usable when creating models. Evolution in the design of airplanes has been deliberately engineered by humans using explicit knowledge and data since just the beginning of the last century. In contrast, for millennia, human evolution has been ongoing, often in unknown ways, in response to our Earth's varied environments, leading to the rich diversity and individuality of the human race. We cannot model what we do not know. This sobering lack of data and knowledge precludes not only a deterministic model of the entire human body, but also the ultimate goals of understanding, treating, or curing all diseases afflicting humankind. Another difference is that airplanes are man-made according to explicit knowledge and data. Tight design, manufacturing, and performance specifications eliminate or limit variability. In contrast, patient variability is an inherent fact of clinical practice. A repeatable simulator without patient variability is, in that sense, an oddity. Repeatable simulator output can be anticipated after repeated use, therefore, making the training exercise less challenging over time.

Variability within a prescribed envelope of simulator output can prevent memorization or recall of prior scenarios from assisting trainees when they subsequently encounter the same scenarios because the time course, dosage, and response of the simulated patient may be different from previous runs. Advocating for variability is not the same as suggesting that simulator output should be unpredictable to the simulator and scenario developers. Within the scope and domain that a scenario is designed to simulate, the output of a simulator, even if it cannot be anticipated through recall or appears unpredictable to trainees, should be predictable to the developers of both the simulator and the scenarios. Unpredictable (to the

developer) simulator output can portray aberrant patient response outside the design envelope, providing undesirable negative teaching.

Patient variability makes a case against simulator repeatability while also fortuitously relaxing modeling requirements: model outputs do not have to match exact values and instead only have to be within a “plausibility” range and within the consensus clinical experience of subject matter experts. Patient variability can be simulated by different techniques. One method implemented for a web-enabled, screen-based propofol simulation at the Virtual Anesthesia Machine (VAM) website⁵ randomizes certain model parameters (such as the volume of distribution) within the range or standard deviation reported in the literature. Another method uses stochastic (a.k.a. probabilistic) models such as the Monte Carlo model. A stochastic simulation of the pharmacokinetics of a two-compartment model⁶ can be viewed at the VAM website. Because of the probabilistic modeling, the design expectation is that the output of a stochastic model will be similar but not repeatable from one run to the next for the same input. Another way to achieve patient variability is to intentionally program different values for given parameters such as race-specific propofol pharmacodynamics where these differences are known; the scenario runs differently based solely on the patient’s race, with everything else being identical, including inputs and interventions.⁷

Patient variability can take place at many levels, including (a) inter-racial variability in average parameters such as race-specific effect site concentration of propofol at loss and recovery of consciousness,⁷ (b) intra-racial, inter-phenotype variability such as redheads being more sensitive to pain and requiring more anesthesia than other Caucasians,⁸ (c) intra-racial, inter-patient, intra-phenotype,⁵ and (d) intra-patient where the parameters of a given patient may change over time such as during pregnancy. Taking the example of inter-racial variability: a simulator that has perfect repeatability when running a propofol-based conscious sedation scenario in countries with predominantly Caucasian populations and also in India would actually be undesirable because, on average, Indian patients are significantly more sensitive to propofol than are Caucasian patients.⁹ Patient variability calls for patient-centered simulation,⁷ which is still in its infancy but has the potential to accelerate the current trend toward adoption and routine practice of patient-centered medicine.

The authors are to be commended for taking a first step in assessing simulator repeatability even though they only assessed intra-simulator repeatability. There seems to be no apparent reason why their work cannot be extended to inter-simulator repeatability that is directly relevant to scenario portability. Further work in simulator repeatability could also address, in those situations where repeatability is relevant, an acceptable repeatability threshold (e.g., is 70%–80% repeatable, instead of 100%, acceptable?) for *clinical training* with mannequin patient simulators integrated with actual monitoring equipment. Repeatability should not be construed as a measure of “performance” of a simulator and indiscriminately demanded as a “must have” property of simulators. Uninformed demands for repeatability run the risk of unnecessarily driving up the costs of simulators and perpetuating the concept of the mythical but so convenient “standard patient” that reinforces the practice of mindless “cookie cutter” medicine. Readers are cautioned that the concept of simulator repeatability must be used

judiciously and should not be indiscriminately required of all simulators or inappropriately applied to “compare apples and oranges.”

Acknowledgments

This work should be attributed to the Department of Anesthesiology, University of Florida College of Medicine, Gainesville, FL; the Center for Safety, Simulation & Advanced Learning Technologies; the University of Florida Clinical & Translational Science Simulation Core and the UF Health Shands Experiential Learning Center. Funded in part by the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH/ NCATS Clinical and Translational Science Award to the University of Florida UL1 TR000064)

References

1. Cumin D, Chen C, Merry AF. Measuring the repeatability of simulated physiology in simulators. *Simul Healthcare*. 2015 In press.
2. Luria I, Lampotang S, Schwab W, Cooper LA, Lizdas D, Gravenstein N. Automated, real-time fresh gas flow recommendations alter isoflurane consumption during the maintenance phase of anesthesia in a simulator-based study. *Anesth Analg*. 2013; 117:1139–1147. [PubMed: 24108253]
3. Garden AL, Robinson BJ, Arancibia CU, Carron TJ, Monk S, Vollmer J, Heinrichs W, Grube C, Graf BM, Johnson EB. Unrecognised malfunction in computerized patient simulators. *Br J Anaesth*. 2004; 93:873–876. [PubMed: 15533959]
4. Sidi A, Gravenstein N, Lampotang S. Construct validity and generalizability of simulation-based objective structured clinical examination scenarios. *J Grad Med Educ*. 2014; 6:489–494. [PubMed: 26279774]
5. Lampotang, S.; Lizdas, D.; Gravenstein, N.; Yavas, S. Web simulation of propofol pharmacokinetics. from University of Florida Department of Anesthesiology Virtual Anesthesia Machine Website; available at: <http://vam.anest.ufl.edu/simulations/propofol.html> [Accessed August 22, 2015]
6. Lampotang, S.; Lizdas, D. Stochastic visualization of a two-compartment PK model. from University of Florida Department of Anesthesiology Virtual Anesthesia Machine Website; available at: <http://vam.anest.ufl.edu/simulations/secondorderstochasticsim2.html> [Accessed August 22, 2015]
7. Jendrusch J, Lampotang S, Lizdas D, Gravenstein N, Ham D, Lok B, Quarles J. Virtual humans for inter-ethnic variability training in sedation and analgesia. *Stud Health Technol Inform*. 2014; 196:175–181. [PubMed: 24732502]
8. Liem EB, Lin CM, Suleman MI, Doufas AG, Gregg RG, Veauthier JM, Loyd G, Sessler DI. Anesthetic requirement is increased in redheads. *Anesthesiology*. 2004; 101:279–283. [PubMed: 15277908]
9. Puri GD, Mathew PJ, Madhavan JS, Hegde HV MBBS MD, Fiehn A. Bi-spectral index, entropy and predicted plasma propofol concentrations with target controlled infusions in Indian patients. *J Clin Monit Comput*. 2011; 25:309–314. [PubMed: 21964767]

FIGURE 1.

A schematic comparison of (top) a physical high-integration mannequin patient simulator connected to actual medical equipment to create a physical simulated environment versus (bottom) a screen-based simulator entirely self-contained within a virtual environment. The difference between a model and a simulator is also depicted.