One Drop at a Time: Research to Advance the Science of Simulation

David A. Cook, MD, MHPE

The article by Weinger[1] in this issue describes a useful framework for thinking about research in simulation. It emphasizes the need for purposeful and programmatic accumulation of information to guide our use of simulation. The analogy with a drug is apt, albeit inexact, given the variations in dose (instructional design and intensity), dosing (repetition), disease (learning objectives), and drug-host, drug-drug, and drug-environment interactions (learner, other learning, and larger context), we encounter in the development of simulation activities. However, although this framework outlines information required to advance our understanding of how to use simulation, it does not spell out the research designs that will be required to accomplish this lofty but laudable objective. In this editorial, I will discuss the types of research needed to accrue the information Weinger requests.

A CRITIQUE OF WEINGER'S FRAMEWORK

However, there are limitations that must be recognized in Weinger's framework. First, simulation is not a drug. One cannot prescribe 20 mg of simulation the way a physician might prescribe 20 mg of lisinopril.[2] Not only does the word simulation encompass a vast range of techniques,[3] but also even within a specific technique (eg, mannequin) and group of learners (eg, medical students) the educational objectives, fidelity, environment, and instructional methods vary widely.[4] The implications for pharmacokinetics and pharmacodynamics studies as proposed by Weinger are obvious: it is difficult to establish a meaningful curve when the drug activity changes each time you study it! Moreover, discussing simulation as a single drug is simplistic; in actuality, we are developing hundreds of drugs to treat hundreds of diseases. The number of permutations in this puzzle is dizzying! These problems are not insurmountable, but they pose a significant challenge.

Second, knowledge/skill deficiency is not a disease. Although it proves useful in this framework, this analogy implies a knowledge/skill transfer model of learning and we know that is not really how people learn.[5,6] Educators and researchers who adopt this framework must remember that teachers and instructional activities facilitate learning, but ultimately it is the learner who must do the work to improve his or her competence.

These limitations are not too serious as long as one does not take the pharmacology model too literally. For example, I am not sure actually creating dose-response curves will be possible, but the concept of a dose-response curve can help us ask important research questions related to intensity and repetition of simulation training. Thus, despite these limitations, I find Weinger's model useful primarily because it highlights the need for multiple types of evidence to clarify our use of simulation.

COMPARATIVE RESEARCH IN SIMULATION

When a new technology is first applied to education, the initial reaction seems to be to test it against no intervention to see whether it works.[7] Thus, we see educators developing learning activities on various topics, all using the new technology, and evaluating performance before and after the simulation-based course or making comparison with no instruction. However, in nearly all cases, the result is similar: if we teach people, they learn.[8–10] The infrequent exceptions to this rule can usually be explained by an inadequate control group (ie, the control group received something more than "no intervention") or inadequate power (sample size too small).

From the Office of Education Research and Division of General Internal Medicine, College of Medicine, Mayo Clinic, Rochester, MN.

Reprints: David A. Cook, MD, MHPE, Division of General Internal Medicine, College of Medicine, Mayo Clinic, Baldwin 4-A, 200 First Street SW, Rochester, MN 55905 (e-mail: cook.david33@mayo.edu).

1

Once educators have begun to demonstrate that the new technology works, the next natural step compares this technology with older educational approaches to see whether the new approach (the new instructional medium) is superior or at least noninferior. Once again, the results are somewhat predictable. If the interventions use similarly effective instructional methods (eg, interactivity, opportunities for application, feedback, and repetition) and similar time learning, there is usually no difference between the media formats. Conversely, if one intervention uses stronger instructional methods or longer time on task, then results typically favor that medium.

However, the limitations of both no intervention-comparison and media-comparative studies go beyond the predictability of the results. Namely, these studies largely fail to inform future practices. It is difficult to take the results of such studies, conducted in a single local setting at a single point in time, and apply (generalize) them to new situations. In a no intervention-comparison study, if (when) we find a significant difference, we have no way of knowing what will happen in a different context or with different learners or whether similar results could be observed using a less intensive (and less expensive) intervention. In a typical media-comparative study (eg, lecture vs. simulator), the instructional methods vary along with the medium, resulting in confounding[11]—more than one explanation for the observed results. Regardless of the study findings, it is impossible to know whether it was the medium or the instructional methods—or both—that accounted for the observation. Assuming a defensible study design, the findings may be true and may serve local purposes (eg, a report to the Dean or funding agency), but it is difficult to take these results and apply them to new courses and contexts because we cannot attribute causality to a specific variation. This does not mean that such studies serve no purpose only that they do not ideally inform future educational activities. In addition, they do not provide the information requested by Weinger.

Weinger suggests that we need to explore issues around the use of simulation presuming it is effective. I believe this is an appropriate approach. The value of simulation to medical training will not be determined by randomized trials, but by logical arguments—informed by evidence—that set forth the values and priorities of society. The question is not "Do we need simulation?" or "Is simulation useful?" The answer to both questions is clearly "Yes!"—for reasons that have been eloquently stated by others[12,13] including patient safety, standardization of training experiences, and performance assessment. Rather, the relevant questions are "when should we use simulation? and how do we effectively use it (eg, what type, how much, and what design) when we do?"[14] These are the questions that need answering, questions that clarify how and why simulation works, and how it can be improved.[15]

RESEARCH TO ADVANCE THE SCIENCE OF SIMULATION

So, if comparisons with no intervention and alternate media are not the answer, what simulation research should we be doing? Broadly, I see at least four classes of research studies. All are important; the order is arbitrary.

First, we need comparative studies but not comparisons with no intervention or with nonsimulation methods of teaching. Rather, we need head-to-head studies comparing one simulation format with another. Studies might compare different levels of simulation fidelity, different combinations of simulation techniques, different instructional techniques, or different sequencing, duration, and repetition of training. Such simulation-simulation comparisons could inform the pharmacokinetic, pharmacodynamic, dose-time, dosing regimen, dose-effect, and drug-drug interaction/synergy relationships requested by Weinger. For example, one study[16] found no difference in epidural catheter performance after training using either a high-fidelity part-task trainer or a low-fidelity model built with a banana and bread.[17] Another study[18] found no difference in vascular anastomosis skill after simulation training with or without videotaped feedback. These studies have important and broadly applicable implications for the design of simulation.

Second, we need association studies. No human has conducted a randomized trial in astronomy, but we have learned a lot about planetary motion through diligent observation. Similarly, we can learn much by exploring relationships among and between aspects of the learning experience that are not amenable to change, such as learner characteristics (eg, experience, spatial ability, and motivation), contextual features (simulation, workplace, and institutional environments), specific educational objectives/content areas, and outcomes such as satisfaction, performance, and noncompulsory usage (eg, after-hours use). Longitudinal studies exploring learning decay and reinforcement could use a similar design. These studies provide data on essentially the same research domains as the comparison studies noted above (pharmacokinetics, pharmacodynamics, dose-time, etc.), but the results will be more tentative, and the hypotheses advanced will generally require confirmation in subsequent research. In one recent association study,[19] investigators found no correlation between performance deterioration between practice sessions and the intersession time interval but found that learners with shorter intersession intervals achieved mastery in fewer sessions.

Third, we need validity studies. We need to know that the measurements in a study actually reflect what they purport to represent (the underlying construct). Construct validity goes far beyond demonstrating an association between learner experience and performance, although this might comprise one important piece of evidence among many. Rather, evidence to support the construct validity of an instrument's scores comes in five flavors[20,21]—content (how well the instrument's items or conditions reflect the intended outcome [construct]), response process (how well actual responses [from learners, assessors, or machines] reflect our intent), internal structure (typically factor analysis or reliability data), relations to other variables (appropriate associations with, for example, a concurrently-administered measure, trainee experience, or future outcomes), and consequences (does our use of this assessment have desired [or undesired] effects?). One such study[22] compared two methods for scoring

a crisis resource management simulation by collecting evidence regarding the content (full representation of content domain), response process (data integrity), internal structure (reliability), and relations to other variables (discrimination among learner experience). Validity studies also might explore the link between intermediate outcomes (knowledge, skills, or simulator-based assessment) and higher order outcomes such as behaviors in practice and effects on patients.[23] If we can firmly establish such links, then we can use more easily measured intermediate outcomes as surrogates in subsequent studies, confident that proximate gains will likely translate to desired real-world benefits. As Weinger notes, we will still need to assess transfer (pharmacodynamics) in select cases, but because pharmacokinetics studies are simpler and cheaper, it makes sense to use these when possible.

Finally, we need rigorous qualitative studies. There is a difference between descriptive studies that report informally derived lessons learned, and studies conceived in a qualitative paradigm and executed using rigorous qualitative methods. Qualitative research is ideally suited to explore the complexity surrounding interpersonal dynamics and contextual factors in simulation activities and can also inform pharmacokinetics and drug-drug interactions. Among other things, qualitative research will help us understand when to use simulation. This information, which complements the quantitative designs described above, will be invaluable. For example, qualitative studies have explored the strengths and weaknesses of standardized patients,[24] key elements in crisis simulation training,[25] and team dynamics in simulated stressful situations.[26]

IN CLOSING

We must do more than to increase the volume and methodological quality of research in simulation. We need to be thoughtful about the research questions we address. Given finite time and resources, we must channel our efforts to those activities that will provide the greatest return on investment. I do not believe that comparisons with no intervention or with nonsimulation instruction provide the highest yield in most cases. Whether we use Weinger's framework or some other agenda,[4,14,27,28] we need research that advances the science[15]—research that helps us understand when to use simulation and how to use it effectively. This clarification research is not as glamorous as comparisons with no intervention, the effect sizes will be smaller (and thus require much larger samples), and advances will typically be incremental rather than revolutionary. Clarification research is less intuitive than studies comparing simulation with traditional methods. The simulation research of which I speak requires relentlessly building on previous work,[27] critically appraising and testing theory,[29,30] and systematically—programmatically[31]—refining our understanding of how to use the tools at our disposal to teach and assess.

It takes years of diligent research to bring a drug to market and additional years to understand how to effectively use the drug alone or in combination to improve the human condition. We should expect nothing less in education science. There are no magic bullets and no shortcuts. Evidence will accumulate slowly, as drops in a bucket. The reward, although delayed, will be worth the work.

REFERENCES

1. Weinger MB. The pharmacology of simulation: a conceptual framework to inform progress in simulation research. *Simul Healthc* 2010;5:8–15.

2. Norman G. RCT = results confounded and trivial: the perils of grand educational experiments. *Med Educ* 2003;37:582–584.

3. Gaba DM. The future vision of simulation in healthcare. *Simul Healthc* 2007;2:126–135.

4. Groom JA. Creating new solutions to the simulation puzzle. *Simul Healthc* 2009;4:131–134.

5. Norman GR, Schmidt HG. The psychological basis of problem-based learning: a review of the evidence. *Acad Med* 1992;67:557–565.

6. Bransford JD, Brown AL, Cocking RR, for the Committee on Developments in the Science of Learning and the Commission on Behavioral and Social Sciences and Education of the National Research Council, editors. *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press; 2000.

7. Cook DA. The failure of e-learning research to inform educational practice, and what we can do about it. *Med Teach* 2009;31:158–162.

8. Clark RE. Reconsidering research on learning from media. *Rev Educ Res* 1983;53:445–459.

9. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. *JAMA* 2008;300:1181–1196.

10. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *Br J Surg* 2008;95:1088–1097.

11. Cook DA. Avoiding confounded comparisons in education research. *Med Educ* 2009;43:102–104.

12. Issenberg SB, McGaghie WC, Hart IR, et al. Simulation technology for health care professional skills training and assessment. *JAMA* 1999; 282:861–866.

13. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med* 2003;78:783–788.

14. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med* 2005;80:541–548.

15. Cook DA, Bordage G, Schmidt HG. Description, justification, and clarification: a framework for classifying the purposes of research in medical education. *Med Educ* 2008;42:128–133.

16. Friedman Z, Siddiqui N, Katznelson R, Devito I, Bould MD, Naik V. Clinical impact of epidural anesthesia simulation on short- and long-term learning curve: high- versus low-fidelity model training. *Reg Anesth Pain Med* 2009;34:229–232.

17. Leighton BL. A greengrocer's model of the epidural space. *Anesthesiology* 1989;70:368–369.

18. Backstein D, Agnidis Z, Sadhu R, MacRae H. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Can J Surg* 2005;48:195–200.

19. Stefanidis D, Walters KC, Mostafavi A, Heniford BT. What is the ideal interval between training sessions during proficiency-based laparoscopic simulator training? *Am J Surg* 2009;197:126–129.

20. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association; 1999.

21. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006; 119:166. e7–e16.

22. Kim J, Neilipovitz D, Cardinal P, Chiu M. A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients

during simulated emergencies (abbreviated as "CRM simulator study IB"). *Simul Healthc* 2009;4:6–16.

23. Aucar JA, Groch NR, Troxel SA, Eubanks SW. A review of surgical simulation with attention to validation methodology. *Surg Laparosc Endosc Percutan Tech* 2005;15:82–89.

24. Bokken L, Rethans JJ, van Heurn L, Duvivier R, Scherpbier A, van der Vleuten C. Students' views on the use of real patients and simulated patients in undergraduate medical education. *Acad Med* 2009;84:958–963.

25. Arora S, Sevdalis N, Nestel D, Tierney T, Woloshynowych M, Kneebone R. Managing intraoperative stress: what do surgeons want from a crisis training program? *Am J Surg* 2009;197:537–543.

26. Weller JM, Janssen AL, Merry AF, Robinson B. Interdisciplinary team interactions: a qualitative study of perceptions of team function in simulated anaesthesia crises. *Med Educ* 2008;42:382–388.

27. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;27:10–28.

28. McGaghie WC, Pugh CM, Wayne DB. Fundamentals of educational research using clinical simulation. In: Kyle R, Murray W, eds. *Clinical Simulation: Operations, Engineering, and Management.* Burlington, MA: Academic Press; 2008:517–526.

29. Bradley P, Postlethwaite K. Simulation in clinical learning. *Med Educ* 2003;37(suppl 1):1–5.

30. Kneebone R. Evaluating clinical simulations for learning procedural skills: a theory-based approach. *Acad Med* 2005;80:549–553.

31. Bordage G. Moving the field forward: going beyond quantitative-qualitative. *Acad Med* 2007;82(suppl 10):S126–S128.