Perspective: **Reconsidering the Focus on "Outcomes Research" in Medical Education: A Cautionary Note**

David A. Cook, MD, MHPE, and Colin P. West, MD, PhD

Abstract

Researchers in medical education have been placing increased emphasis on "outcomes research," or the observable impact of educational interventions on patient care. However, although patient outcomes are obviously important, they should not be the sole focus of attention in medical education research. The purpose of this perspective is both to highlight the limitations of outcomes research in medical education and to offer suggestions to facilitate a proper balance between learner-centered and patient-centered assessments. The authors cite five challenges to research using patient outcomes in medical education, namely (1) dilution (the progressively attenuated impact of education as filtered through other health care providers and systems), (2) inadequate sample size, (3) failure to establish a causal link, (4) potentially biased outcome selection, and (5) teaching to the test. Additionally, nonpatient outcomes continue to hold value, particularly in theory-building research and in the evaluation of program implementation. To educators selecting outcomes and instruments in medical education research, the authors offer suggestions including to clarify the study objective and conceptual framework before selecting outcomes, and to consider the development and use of behavioral and other intermediary outcomes. Deliberately weighing the available options will facilitate informed choices during the design of research that, in turn, informs the art and science of medical education.

Over the past decade, the academic community has seen increased emphasis on "outcomes research" in medical education.[1–3] Although the word *outcome* refers generally to "something that follows as a result or consequence,"[4] in the discourse of outcomes research in medical education, this word refers to clinical outcomes—that is, an intervention's impact on patients and sometimes on physician behaviors during patient care.[2] This new emphasis on outcomes is laudable because the ultimate intent of medical education is to improve the health of patients. Focusing on what really works helps to promote responsiveness to social priorities, highlights inefficient and ineffective education practices, and encourages attention to care systems. Although authors have reported patient outcomes in only a small minority of medical education studies thus far,[1,5–7] the number of such studies appears to be growing as evidenced by their prevalence in recent systematic reviews.[8–10] Investigators have used patient outcomes in studies of professional behavior,[11] physician communication,[12] surgical training,[13] and continuing medical education.[14] Others have reported systems to collect patient outcomes longitudinally in postgraduate education.[15,16]

Dr. Cook is professor of medicine and medical education and director, Office of Education Research, College of Medicine, Mayo Clinic, Rochester, Minnesota.

Dr. West is associate professor of medicine and biostatistics and associate director, Internal Medicine Residency Program, College of Medicine, Mayo Clinic, Rochester, Minnesota.

Correspondence should be addressed to Dr. Cook, Division of General Internal Medicine, Mayo Clinic, 200 First St. SW, Rochester, MN 55905; e-mail: cook.david33@mayo.edu.

Acad Med. 2013;88:162–167.
First published online December 23, 2012
doi: 10.1097/ACM.0b013e31827c3d78

Although the patient outcomes research movement is important, patient effects and physician behaviors make up only a subset of possible outcomes. Fifty years ago, Kirkpatrick[17] proposed a widely accepted, four-level model of training program outcomes, comprising, first, reaction (satisfaction); followed by learning (knowledge, skills, and attitudes); then, behaviors in practice; and, finally, results (effects on the object of interest, such as in medicine, patients). Patient outcomes warrant emphasis, but they should not constitute the sole focus of attention in medical education.

An excessive emphasis on patient outcomes may paradoxically distract investigators from advancing the art and science of medical education overall. Such an emphasis on patient outcomes in medical education would be akin to focusing clinical research outcomes on mortality, which would neglect other outcomes important to patients (such as quality of life), restrict the type of research questions asked (not all interventions are designed to prolong life), and make many studies infeasible (e.g., randomized trials with mortality outcomes typically require long periods of follow-up and incur high expense).

We recognize several drawbacks to using patient outcomes in medical education research. Although we acknowledge that all of these limitations do not apply to every situation and that none are insurmountable, we believe that, collectively, they present a formidable challenge. The purpose of this perspective is both to highlight the limitations of patient outcomes research in medical education and to offer suggestions to facilitate a proper balance between learner-centered and patient-centered assessments in medical education research. We seek not to discourage research involving patient-oriented outcomes but simply to counterbalance calls for whole-scale adoption of the patient outcomes perspective as the holy grail of education research.

Challenges and Limitations of Research Using Patient Outcomes in Medical Education

Dilution

The link between what a physician does (Kirkpatrick's *behaviors*) and what

patient outcomes reflect (Kirkpatrick's *results*) is indirect. A physician's actions mingle with patients' preferences for therapy or testing, patients' compliance with physician recommendations, and individual variations in disease and demographics.[18] Moreover, in most instances, additional factors—such as other members of the health care team (nurses, pharmacists, trainees, etc.), institutional policies, and insurance plan requirements—also have an effect. Each of these factors could correct undesirable physician behaviors, or, alternatively, they could effectively hide correct behaviors from manifesting in detectable change. Ultimately, these confounding factors or conditions effectively dilute the physician's actions[19] and diminish the observed effect of the educational activities that preceded those actions. Shifting from practicing physicians to physician trainees (residents) or medical students adds further levels of dilution that only magnify the problem. Researchers have two options: to increase the initial impact of the intervention (so that even after dilution the effect remains strong) or to enroll a sample size large enough to detect even small (dilute) effects. Neither solution is ideal in education research.

Creating an exceptionally strong intervention sounds attractive at first. However, we have observed that strong interventions nearly always require a multifaceted approach to training, drawing on multiple instructional modalities (e.g., combinations of textbook, video, lecture, small groups, computer-assisted instruction, standardized patients, other simulation, and clinical encounters) and instructional methods (practice cases, group discussion, self-assessment questions, feedback, mastery learning, etc.). One of us and another colleague[20] previously noted

> When complex interventions show significant benefit, they demonstrate that a specific outcome (e.g. knowledge or behavior) can be modified but tell us little about which components of the intervention (e.g. instructional methods and experiences) determined this change. Such investigations have only limited generalizability because the multifactorial intervention cannot be replicated precisely, and implementing only a portion of the intervention may or may not be effective.

Moreover, strong interventions often show a large effect only when judged against a weak comparison intervention or no intervention. As researchers seek to advance the science of education, the importance of comparative effectiveness research (side-by-side comparisons of two or more active educational interventions) will increase.[21] The expected effect size in such research is often rather small.[22–24]

Some have argued that the community might shift its attention to assessing practitioner groups, such as looking at the aggregate data for a training program[15] or large cohort.[25] Although examining aggregate group data makes sense at a programmatic level (i.e., for identifying curricular priorities and gaps, or for demonstrating the overall effectiveness of a program), doing so minimizes the importance of individual providers. Although many health care problems require systems-based or systems-level solutions to achieve demonstrable and sustained change, it is still the individual who graduates from medical school, qualifies for and maintains a license and board certification, and—in most instances—sits in the room with a patient during the clinical encounter.

Feasibility: Sample size

In medical education research, the conveniently available sample size (e.g., the number of participants in a training program) is often inadequate to appropriately power the study. Clinical trials frequently enroll thousands of patients to evaluate the effectiveness of therapeutic options. Consider how many physicians (let alone medical students) investigators would need to enroll in order to study the patient outcomes effect of teaching physicians about the benefits of metoprolol succinate in heart failure, ramipril in intermediate-risk patients, or clopidogrel in stroke prevention—benefits demonstrated in very large clinical trials.[26–28] Because the effect of the educational intervention—even if successful—would be diluted (as above), and the measurements would be imperfect, such a study would either need a very large sample or an intervention with a huge impact (large effect size). Anything less would likely result in nonstatistically significant findings. For example, to demonstrate an association between certification exam scores and patient complaints, one study enrolled all 3,424 physicians certified in the provinces of Quebec and Ontario during a four-year period.[12] (In an earlier and somewhat smaller study of 934 physicians, the same authors demonstrated a link between exam scores and physician *behaviors* such as ordering tests and prescribing.)[29]

Some investigators have attempted to overcome the barrier of an insufficient *learner* sample by enrolling more *patients*. However, they must then account for clustering when analyzing the data, and clustering lowers the effective sample size.[30] Regrettably, researchers often fail to adjust for clustering (as documented in clinical research),[31] resulting in flawed analyses and questionable interpretations, as noted in recent systematic reviews in education.[8,32]

Other researchers have attempted to increase sample size by enrolling learners from multiple programs—either different training programs within an institution or similar programs from different institutions. Although doing so often leads to success, many important research questions do not lend themselves to multiprogram study—especially questions that require interventions and outcome measures to be implemented equally across programs.[33]

Failure to establish a causal link

Educators might expect a focus on patient outcomes to improve study rigor. Regrettably, studies using patient outcomes often suffer from threats to both internal study validity (the absence of bias in the study findings) and external study validity (the meaningfulness of the findings to others). Such threats to validity undermine the inferences and conclusions drawn by investigators and readers. Many medical education studies using patient outcomes employ nonrandomized designs such as concurrent cohort designs, retrospective designs with historical controls, and single-group cross-sectional designs. Such designs allow much weaker causal interpretations than do randomized studies.[34]

Another source of weakened causal inferences is confounding, which occurs when the observed effect can plausibly be ascribed to a known or suspected

influence other than the object under study. Multifaceted interventions (as described above in the Dilution section) and comparisons in which multiple instructional features vary simultaneously represent two common sources of confounding in education research.[35] A third source of confounding, particularly problematic for patient outcomes research, is systematic variation in patient populations across physicians (e.g., some physicians care for higher-risk patients than others).[18] Even randomization cannot compensate for a confounded design.[20]

The tensions between outcomes and other aspects of study design raise the following question: Which is better, a study that permits strong causal interpretations (e.g., strong design and limited confounding) with a weak outcome, or a study with a strong outcome but a design that allows for only weak and confounded interpretations? The answer depends on the situation, but in many cases the stronger causal design may be preferable. Being able to state that outcomes improved, without establishing a clear link to what actually caused that improvement, does little to advance the community's understanding of how to enhance future clinical/educational practice.[35]

Potentially biased outcome selection

The following anecdote illustrates a fourth limitation of patient outcomes research in medical education:

> A woman walking down the street one night noticed a man on his knees under a lamppost. When asked what he was doing, the man replied that he was looking for his keys. She joined him in the search, but after several minutes asked, "Are you sure you lost them here?" "Oh no," the man replied. "I dropped them on the other side of the street. But it's dark over there; the light is much better here."

When researchers assess patient-level outcomes, they would ideally assess the outcomes of greatest significance. Yet, seemingly, they often search where they find light rather than where they lost the keys. For example, researchers conducting a recent systematic review of simulation-based education found that all of the studies reporting patient outcomes focused on procedural tasks (e.g., endoscopy and endotracheal intubation),[32] whereas no studies used patient outcomes to evaluate simulation-based training activities for no-less-important nonprocedural tasks (e.g., physical exam or crisis resource management).[8]

Researchers risk bias when they select an outcome that does not reflect the entire domain of interest. Regrettably, many of the relevant outcomes in education research do not readily lend themselves to measurement,[18] and researchers thus select measures that are easy rather than those that best reflect broad curricular goals. Many important clinical activities have no accepted standard,[18] making the corresponding clinical metrics impossible to use as research outcomes. For example, the optimal frequency of bone density screening remains unclear, making screening frequency inadequate as a measure of curricular impact in the topic of osteopenia. Some conditions (i.e., topics) also lend themselves to patient outcome assessment more readily than others. Checking a lab test result (e.g., hemoglobin A1c) is easier than reliably determining the rate of smoking cessation. Moreover, even for a given condition, some measures (e.g., hemoglobin A1c) are easier to quantify than others (e.g., onset of diabetic peripheral neuropathy). Medical education researchers often lack funding to prospectively monitor patient outcomes; therefore, they select measures available from the medical record. Although easy to collect, such data may not be good indicators of actual performance.[18] Selecting a given outcome for reasons of availability or feasibility introduces possible bias in the clinical topic or the measurement approach.

Looking under the lamppost is not always a problem. Researchers wishing to demonstrate "proof of concept" might reasonably select a test-case clinical topic with a patient outcome that is intentionally easy to measure. However, as authors have noted for both clinical outcomes[18] and assessment in general,[36] performance in one domain often has little correlation with another. For example, superior performance in managing blood glucose in diabetes may not predict performance in colon cancer screening. As the field moves beyond proof of concept, the continued selection of easy-to-measure outcomes (topics) at the exclusion of more difficult but equally important outcomes (topics) risks unjustifiable bias. Those engaged in patient outcomes research must ensure that the sampling of topics adequately reflects the entire curriculum.

Teaching to the test

Conventional wisdom indicates that assessment drives learning, and most educators (including ourselves) would agree that this is usually a good thing. Assessment can motivate and focus both learners and teachers to address learning gaps they might otherwise overlook. However, excessive focus on patient-oriented outcomes could negatively affect teaching by leading educators to "teach to the test."

Too much attention to patient outcomes could lead curriculum designers to teach only those processes that unambiguously enhance patient care. Although seemingly sensible, this approach suffers from at least two shortcomings. First, despite the research community's valiant attempts to improve the situation, clear evidence informs only a fraction of clinicians' diagnostic and therapeutic decisions.[37] Focusing primarily on practices with defined standards will necessarily detract from teaching on other topics. Second, focusing on evidence-based algorithmic approaches to management could backfire if learners fail to learn the principles that underlie such actions. Although systems change usually has a stronger effect on patient outcomes than education,[38] learning pathophysiology and other underlying principles has clear benefit on retention and transfer,[39] to say nothing of the long-term benefits of such knowledge[40] in understanding new therapies, interpreting new study results, or conducting research later in life.

We do not wish to be misinterpreted. Clearly, educators should teach and reinforce clinical actions that benefit patients. However, many vital activities will not have an immediate, visible impact on patients. Focusing excessively on improving measurable patient outcomes could lead to short-term gains and long-term losses.

Patient Outcomes: Not Always Better

The argument that patient outcomes are superior to other outcomes is, ultimately, a value judgment. Shea[19] pointed out that "the primary customer of medical

education is emphatically the learner, not the patient." Of course, prudent physician-in-training customers will want the best value in training that will enable them to provide the most effective patient care. However, measures such as knowledge, skills, attitudes, time, and even satisfaction should not automatically be relegated to second-tier status as "process measures."[2] All else being equal, higher learner satisfaction is not a bad objective! Also, as Yardley and Dornan[9] have noted, the community can learn much about an educational activity from outcomes lower in Kirkpatrick's model and from nonoutcomes evidence (e.g., process evaluation and qualitative data). Furthermore, excessive concentration on patient outcomes risks dehumanizing trainees (and thereby the education process) by viewing the trainee solely as a means to an end rather than a worthy end in and of him- or herself.

Nonpatient outcomes (knowledge and skills) may be particularly important in theory-building research because this type of research often occurs in settings with limited patient contact. Medical education researchers have frequently lamented the absence of theory in the field,[41,42] and some have suggested theory-building research as central to advancing the community's understanding of how to improve learning activities.[43] Patient outcomes research may tell the medical education community whether or not something worked, but it often does not clarify how to improve a course for the next go-round or how to effectively design a new course. Only sound theories and conceptual frameworks will permit these advances.

Recommendations

An excessive focus on patient-oriented outcomes will at best distract the education community from important research using other outcomes, and at worst it could adversely affect some aspects of health professions education. Research focused on patient-level outcomes is and will remain essential to evaluate medical education activities, but it should not be pursued at the expense of research using other outcomes. Thus, we offer six suggestions to guide the selection and analysis of outcomes and instruments in medical education research.

First, rather than starting a research project by identifying a measure or tool (e.g., "hemoglobin A1c" or "the patient record") and then designing the investigation around it, researchers should first clarify the study objective and conceptual framework, then select the most relevant outcome, then the measurement method, and finally the instrument.[44] By selecting the question first, they both maintain focus on the most important issues and avoid prematurely selecting an outcome or instrument that will not provide the most meaningful data. Researchers must also ensure that the outcomes align with the educational objectives. No "most important" outcome exists in absolute terms—only better outcomes for a given context and purpose. The best outcome will balance two (at times opposing) requirements: the need to provide meaningful conclusions for the intended audience and the constraints of feasibility.

Second, for purposes of clarity in discussing the patient-related outcomes of health professions education, educators should remember the distinction between *skills* (provider actions in an artificial test setting), *behaviors* (provider actions with real patients, such as ordering tests, prescribing, procedural time, or procedural technique), and *patient effects* (Kirkpatrick's level 4 "results": the actual impact on patients, such as patient satisfaction, patient compliance, symptom control, complications, or test results).[8,32] Of note, a patient characteristic such as motivation to change might be considered an attitude in *clinical research*, but we argue that in *health professions education research* this characteristic qualifies as a true patient effect.

Third, researchers need to focus on establishing links between patient outcomes and other more accessible outcomes. To link patient outcomes causally to an educationally relevant activity or personal characteristic can be challenging. If the conceptual relationship between the activity and the outcome is poorly defined, investigators will be unable to bridge the gap with a single link. In such instances, they may find that using two or more links provides a more readily accessible chain of causality. For example, if researchers demonstrate an association between specific skills or behaviors and specific patient outcomes, then they and others may use these skills or behaviors as surrogate outcomes in subsequent studies (see Figure 1). For example, in a simulation-based course on vascular surgery, investigators found that simulator outcomes of time and severity of anastomotic leaks (skills) were associated with operative time and anastomotic leaks in real patients.[45] Another study found an association between the quality of counseling with real patients (a behavior) and the patients' motivation to change (a patient effect).[46] Of course, surrogate outcomes can be misleading,[47] as is well understood in clinical research.[48] Adapting existing guidelines for the use of surrogate end points in clinical research to medical

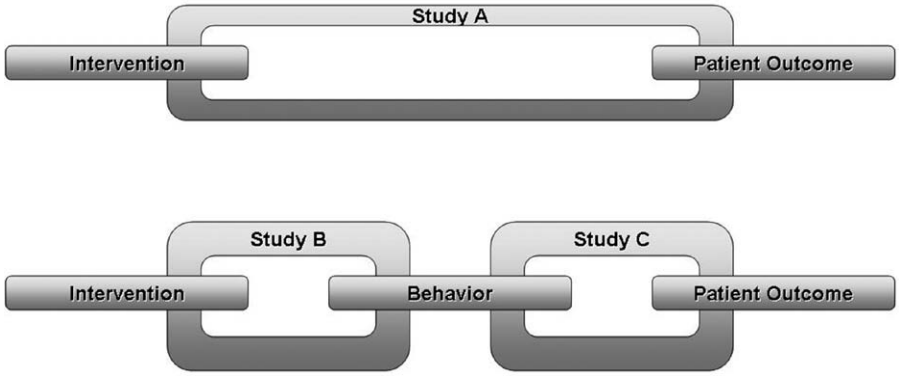

Figure 1 Establishing links from education to patient outcomes. Establishing a direct link between an educational intervention and a meaningful patient outcome often proves difficult. Bridging the gap frequently comes at the expense of being able to make strong causal inferences. For example, Study A might be observational (rather than experimental) or use a multifeatured intervention with limited generalizability. By contrast, both Study B and Study C could use stronger causal designs that provide a chain of evidence that, although indirect, may permit more defensible inferences.

education research seems prudent,[49] including not only that the surrogate must correlate with the patient outcome but that *improvement* in the surrogate should also associate with *improved* patient outcomes.

Fourth, investigators should consider proceeding in a deliberately stepwise fashion as they test educational interventions: first assessing knowledge and skills, then behaviors, and finally patient outcomes. As Shea[19] stated, "Before we shift our focus—and simultaneously the expectations of reviewers and editors—we need to make sure we can influence students' behaviors. Once we know how to do this, we can turn our attention to the next link." Researchers should also consider the learner's training level: The link between behaviors and patient outcomes is much more direct (less diluted) for physicians, and to a lesser extent for postgraduate trainees, than it is for medical students. A study of cardiac resuscitation training for internal medicine residents illustrates the stepwise progression of outcomes: First, the investigators established that the course improved resuscitation skills in a simulated setting[50]; then, in a subsequent study, they assessed behaviors (checklist score during actual resuscitation) and patient outcomes (survival to discharge).[51] Another program of research began with an assessment of the need for training in obesity counseling,[52] proceeded with a study evaluating the impact of training on patient counseling activities (behaviors),[53] and then evaluated the effect on weight change[54] (a patient outcome).

Fifth, investigators might consider selecting patient outcomes that result from the engagement of patients and the whole health care team. Kalet and colleagues[16] have offered a conceptual framework for "educationally sensitive patient outcomes" that focuses on the capacity of individual providers to influence patient care by enhancing patients' active involvement in their own care and by effectively engaging the health care team and available systems. These outcomes (e.g., patient motivation to change or team function) lie at the interface between behaviors and patient outcomes. In addition, they may offer a feasible approach to studies of educational programs that yield insight into patient care effects—provided educators can develop and implement appropriate measurement tools. The study cited above that linked physician counseling and patient motivation[46] illustrates one application of educationally sensitive patient outcomes.

Finally, we remind researchers that advanced statistical techniques will be required whenever there is more than one patient outcome per trainee (i.e., clustering of patients).[30,32] Failure to adjust for clustering when required constitutes a unit-of-analysis error that artificially inflates the study power and may lead to spurious conclusions.

In Sum

Patient outcomes in medical education research have many advantages, but they typically carry some risks as well. Issues such as dilution, feasibility, failure to establish a causal link, potentially biased outcome selection, and teaching to the test all challenge the routine use of patient outcomes. Moreover, they are not the only important outcomes in medical education. Deliberately weighing the available options will facilitate informed choices during the design of research that, in turn, informs the art and science of medical education.

Funding/Support: None.

Other disclosures: None.

Ethical approval: Not required.

References

1 Prystowsky JB, Bordage G. An outcomes research perspective on medical education: The predominance of trainee assessment and satisfaction. Med Educ. 2001;35:331–336.
2 Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. Acad Med. 2004;79:955–960.
3 Dauphinee WD. Educators must consider patient outcomes when assessing the impact of clinical training. Med Educ. 2012;46:13–20.
4 Merriam-Webster online. Outcome [definition]. http://www.merriam-webster.com/dictionary/outcome. Accessed September 17, 2012.
5 Cook DA, Levinson AJ, Garside S. Method and reporting quality in health professions education research: A systematic review. Med Educ. 2011;45:227–238.
6 Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: Quality of submissions to JGIM's medical education special issue. J Gen Intern Med. 2008;23:903–907.
7 Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA. 2007;298:1002–1009.
8 Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. JAMA. 2011;306:978–988.
9 Yardley S, Dornan T. Kirkpatrick's levels and education 'evidence.' Med Educ. 2012;46:97–106.
10 Fletcher KE, Reed DA, Arora VM. Patient safety, resident education and resident well-being following implementation of the 2003 ACGME duty hour rules. J Gen Intern Med. 2011;26:907–919.
11 Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. N Engl J Med. 2005;353:2673–2682.
12 Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. JAMA. 2007;298:993–1001.
13 Prystowsky JB, Bordage G, Feinglass JM. Patient outcomes for segmental colon resection according to surgeon's training, certification, and experience. Surgery. 2002;132:663–670.
14 Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: Do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? JAMA. 1999;282:867–874.
15 Haan CK, Edwards FH, Poole B, Godley M, Genuardi FJ, Zenni EA. A model to begin to use clinical outcomes in medical education. Acad Med. 2008;83:574–580.
16 Kalet AL, Gillespie CC, Schwartz MD, et al. New measures to establish the evidence base for medical education: Identifying educationally sensitive patient outcomes. Acad Med. 2010;85:844–851.
17 Kirkpatrick DL. Techniques for evaluating training programs. J Am Soc Train Dir. 1959;13:3–9.
18 Landon BE, Normand SL, Blumenthal D, Daley J. Physician clinical performance assessment: Prospects and barriers. JAMA. 2003;290:1183–1189.
19 Shea JA. Mind the gap: Some reasons why medical education research is different from health services research. Med Educ. 2001;35:319–320.
20 Cook DA, Beckman TJ. Reflections on experimental research in medical education. Adv Health Sci Educ Theory Pract. 2010;15:455–464.
21 Cook DA. If you teach them, they will learn: Why medical education needs comparative effectiveness research. Adv Health Sci Educ Theory Pract. 2012;17:305–310.
22 Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: A systematic review and meta-analysis. Acad Med. 2010;85:1589–1602.
23 Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in Internet-based learning for health professions education: A

systematic review and meta-analysis. Acad Med. 2010;85:909–922.
24 Cook DA, Hamstra SJ, Brydges R, et al. Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. Med Teach. September 3, 2012. doi:10.3109/0142159X.2012.714886.
25 Carney PA, Nierenberg DW, Pipas CF, Brooks WB, Stukel TA, Keller AM. Educational epidemiology: Applying population-based design and analytic approaches to study medical education. JAMA. 2004;292:1044–1050.
26 Hjalmarson A, Goldstein S, Fagerberg B, et al. Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: The metoprolol CR/XL randomized intervention trial in congestive heart failure (MERIT-HF). MERIT-HF Study Group. JAMA. 2000;283:1295–1302.
27 Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med. 2000;342:145–153.
28 CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). Lancet. 1996;348:1329–1339.
29 Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. JAMA. 2002;288:3019–3026.
30 Kerry SM, Bland JM. Analysis of a trial randomised in clusters (statistics notes). BMJ. 1998;316:54.
31 Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behavior. J Gen Intern Med. 1992;7:623–629.
32 Zendejas B, Brydges R, Wang AT, Cook DA. Patient outcomes in simulation-based medical education: A systematic review. J Gen Intern Med. In press.
33 Cook DA, Andriole DA, Durning SJ, Roberts NK, Triola MM. Longitudinal research databases in medical education: Facilitating the study of educational outcomes over time and across institutions. Acad Med. 2010;85:1340–1346.
34 Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing Clinical Research: An Epidemiologic Approach. 2nd ed. Philadelphia, Pa: Lippincott Williams & Wilkins; 2001.
35 Cook DA. Avoiding confounded comparisons in education research. Med Educ. 2009;43:102–104.
36 Norman GR. The glass is a little full—of something: Revisiting the issue of content specificity of problem solving. Med Educ. 2008;42:549–551.
37 Institute of Medicine. Learning What Works Best: The Nation's Need for Evidence on Comparative Effectiveness in Health Care. http://www.iom.edu/~/media/Files/Activity%20Files/Quality/VSRT/ComparativeEffectivenessWhitePaperESF.pdf. Accessed October 17, 2012.
38 Gilbody S, Whitty P, Grimshaw J, Thomas R. Educational and organizational interventions to improve the management of depression in primary care: A systematic review. JAMA. 2003;289:3145–3151.
39 Woods NN. Science is fundamental: The role of biomedical knowledge in clinical reasoning. Med Educ. 2007;41:1173–1177.
40 Schmidt HG, Rikers RM. How expertise develops in medicine: Knowledge encapsulation and illness script formation. Med Educ. 2007;41:1133–1139.
41 Cook DA, Beckman TJ, Bordage G. Quality of reporting of experimental studies in medical education: A systematic review. Med Educ. 2007;41:737–745.
42 Prideaux D, Bligh J. Research in medical education: Asking the right questions. Med Educ. 2002;36:1114–1115.
43 Bordage G. Conceptual frameworks to illuminate and magnify. Med Educ. 2009;43:312–319.
44 Cook DA. Twelve tips for evaluating educational programs. Med Teach. 2010;32:296–301.
45 Wilasrusmee C, Lertsithichai P, Kittur DS. Vascular anastomosis model: Relation between competency in a laboratory-based model and surgical competency. Eur J Vasc Endovasc Surg. 2007;34:405–410.
46 Jay M, Gillespie C, Schlair S, Sherman S, Kalet A. Physicians' use of the 5As in counseling obese patients: Is the quality of counseling associated with patients' motivation and intention to lose weight? BMC Health Serv Res. 2010;10:159.
47 Boden WE, Probstfield JL, Anderson T, et al; AIM-HIGH Investigators. Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. N Engl J Med. 2011;365:2255–2267.
48 Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011;343:d7995.
49 Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. JAMA. 1999;282:771–778.
50 Wayne DB, Butter J, Siddall VJ, et al. Simulation-based training of internal medicine residents in advanced cardiac life support protocols: A randomized trial. Teach Learn Med. 2005;17:210–216.
51 Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie WC. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: A case–control study. Chest. 2008;133:56–61.
52 Jay M, Gillespie C, Ark T, et al. Do internists, pediatricians, and psychiatrists feel competent in obesity care? Using a needs assessment to drive curriculum design. J Gen Intern Med. 2008;23:1066–1070.
53 Jay M, Schlair S, Caldwell R, Kalet A, Sherman S, Gillespie C. From the patient's perspective: The impact of training on resident physician's obesity counseling. J Gen Intern Med. 2010;25:415–422.
54 Jay M, Gillespie C, Schlair S, et al. The impact of primary care resident physician training on patient weight loss at 12 months. Obesity (Silver Spring). May 25, 2012. doi:10.1038/oby.2012.137.