Limitations of Medical Research and Evidence at the Patient-Clinician Encounter Scale

Alan H. Morris, MD, FCCP; and John P. A. Ioannidis, MD, DSc

We explore some philosophical and scientific underpinnings of clinical research and evidence at the patient-clinician encounter scale. Insufficient evidence and a common failure to use replicable and sound research methods limit us. Both patients and health care may be, in part, complex nonlinear chaotic systems, and predicting their outcomes is a challenge. When trustworthy (credible) evidence is lacking, making correct clinical choices is often a low-probability exercise. Thus, human (clinician) error and consequent injury to patients appear inevitable. Individual clinician decision-makers operate under the philosophical influence of Adam Smith's "invisible hand" with resulting optimism that they will eventually make the right choices and cause health benefits. The presumption of an effective "invisible hand" operating in health-care delivery has supported a model in which individual clinicians struggle to practice medicine, as they see fit based on their own intuitions and preferences (and biases) despite the obvious complexity, errors, noise, and lack of evidence pervading the system. Not surprisingly, the "invisible hand" does not appear to produce the desired community health benefits. Obtaining a benefit at the patient-clinician encounter scale requires human (clinician) behavior modification. We believe that serious rethinking and restructuring of the clinical research and care delivery systems is necessary to assure the profession and the public that we continue to do more good than harm. We need to evaluate whether, and how, detailed decision-support tools may enable reproducible clinician behavior and beneficial use of evidence.

CHEST 2013; 143(4):1127–1135

Obtaining favorable outcomes is a long-standing goal of medicine and health care, tightly linked to the Hippocratic principles of beneficence (to do good-ὠφελείν) and nonmaleficence (to do no harm-μὴ βλάπτειν).^{1,2} Beneficence can either mean producing

benefit (a utilitarian viewpoint, according to Mill) or simply intending to produce benefit (a deontologic viewpoint, according to Kant).³ Utilitarianism is defensible in spite of critiques.^{4,5} Patients want to benefit at the end of the day.⁶

Intentions to do good are widespread but frequently inadequate for producing benefit, as suggested by the well-known aphorism, "The road to hell is paved with good intentions." Besides, current health care involves many influential corporate and other stakeholders whose prime motive and intentions are not beneficent, but rather are monetary. Even physicians and health researchers themselves may be driven by less lofty intentions like financial gains or academic prestige. Regardless, reasonable people would prefer benefit more than just good intentions in the patient-clinician encounter.

Manuscript received July 30, 2012; revision accepted September 1, 2012.

Affiliations: From the Pulmonary and Critical Care Divisions, Departments of Medicine (Dr Morris), Intermountain Medical Center, Intermountain Healthcare and The University of Utah School of Medicine, Salt Lake City, UT; and Stanford Prevention Research Center (Dr Ioannidis), Department of Medicine, and Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA.

Funding/Support: This work was supported by the National Institutes of Health [RO1-HL-36787, NO1-HR-46062], the Agency for Healthcare Research and Quality (HS 06594), the Intermountain Research & Medical Foundation, the National Respiratory Distress Syndrome Foundation, the LDS Hospital, and Intermountain Healthcare, Inc (all to Dr Morris). The work of Dr Ioannidis was supported by an unrestricted gift by Sue and Robert O'Donnell.

Correspondence to: Alan H. Morris, MD, FCCP, Pulmonary/Critical Care Division, Sorenson Heart & Lung Center, 6th Floor, Intermountain Medical Center, 5121 S Cottonwood St, Murray, UT 84157-7000; e-mail: alan.morris@imail.org

© 2013 American College of Chest Physicians. Reproduction of this article is prohibited without written permission from the American College of Chest Physicians. See online for more details.
DOI: 10.1378/chest.12-1908

There are two broad categories of response to these Hippocratic principles.¹ The first is provision of compassion. The second is consistent application of scientific results proved to confer more clinical benefit than harm (a favorable outcome).

Compassion is always possible in the patient-clinician encounter. However, we do not always have interventions documented to provide more benefit than harm. Interventions may be unavailable, their harms may exceed benefits, or evidence may incorrectly suggest more benefit than harm because study results reflect biases and are not trustworthy.^{7,8} If so, offering compassion is worthy, while offering anything more would be detrimental to the patient. Both seasoned and young physicians are often tempted to do more. This includes using unjustified tests (eg, prostate specific antigen screening) or giving untested or poorly tested therapies (eg, antiarrhythmics to suppress asymptomatic arrhythmia).⁹ We often do not want to accept that there is nothing to do and that medicine is not omnipotent. It would be useful to resist these temptations. Medicine is not omnipotent.

TRUSTWORTHY FAVORABLE EFFECTS

Some outcome predictability is a prerequisite for favorable impact in medicine and health care: Based on some prior knowledge of intervention effects and risk factors, we can expect the effect in newly treated patients. Interestingly, predictability does not require full understanding of causal mechanisms. Our inability to fully understand causality is not an unsurpassable barrier.¹⁰ However, outcome predictability requires trustworthy scientific results. Trustworthy or credible results require robust study methods that separate true effect from noise, and reliably estimate effect size. Science requires this true effect to be reproducible,^{11,12} although reproducibility is sometimes difficult to establish, or is absent in medicine, and in other scientific domains.^{13,14}

When we probe too many hypotheses in data-rich settings or when beneficial effects are small, it may be more difficult to separate true signals from noise. Both of these situations seem common in modern medicine. Data-rich electronic health records, administrative databases, and biobanks are assuming increasingly prominent roles. There is great hope, but also hype, that nonrandomized data can provide reliable effect sizes, for associations and for treatment effects. However, it is increasingly clear that most efficacious medical interventions have small effects,¹⁵ even under optimal circumstances. The large effects touted in major medical journals presenting new promising treatments may be grossly exaggerated.^{7,13,16} Even with perfect methods, small effects are difficult to identify and measure with precision. When methods break

down and are applied in suboptimal fashion due to lack of expertise, conflicts, or variable clinician decision-making due to absence of adequate decision-support tools, study results are usually not trustworthy. Consequently, the patient-clinician encounter is surrounded by noise and biased results. Strikingly, the credibility rates even of meta-analyses of randomized trials (the highest level in most “hierarchies of evidence”), and even of those done by an outstanding group without conflicts of interest (Cochrane Collaboration), may be 80% at best, but may be below 50%.¹⁷ For single studies, scattered evidence and nonrandomized designs, the credibility is likely to be much lower.⁷ The Institute of Medicine has recently highlighted the need for clarification of effectiveness of different therapeutic choices, and for improving the validity and utility of guidelines and other comparative effectiveness tools.¹⁸⁻²⁰

HEALTH CARE INVOLVES COMPLEX, NONLINEAR, CHAOTIC SYSTEMS

Both patients and health care in general may be, in part, complex nonlinear chaotic systems with predictive relationships (coefficients of variables, or effects of treatments) that are inconstant.²¹⁻²³ The “butterfly effect” on weather prediction is the most famous example of a nonlinear chaotic system effect. Predicting outcomes in such systems is a challenge, especially when deviating from original conditions of study. When trustworthy (credible) evidence is lacking, the situation is even worse: Making correct choices is often a low-probability exercise. In the absence of credible evidence, it is likely that there are available many more wrong responses than correct ones.²⁴ This reflects the Aristotelian proposition that there are many ways to be wrong and only one way to be correct. Clinician decision-making, when unsupported by credible outcome data, is likely not only to be variable but also to be incorrect. For a given clinical dilemma, there can be many wrong choices.

For both clinical experiments (clinical trials) and clinical care, we do not possess tools to standardize many clinician decisions that include a sequence of choices. Cointerventions (confounders introduced after subject randomization) are not commonly controlled in clinical trials. As a result, clinical trials, and especially nonblinded ones or those where masking is not efficient, may suffer from excess variation and nonreproducible experimental methods. Cointerventions can, therefore, figure among many sources of variable results.²⁵⁻³¹ Unbalanced cointervention delivery due to nonreproducible methods may explain some conflicting clinical trial results observed in different studies of the same putative intervention.^{32,33}

Given this complexity, human error and consequent injury to patients appear inevitable.³⁴⁻³⁷ The Institute of Medicine (National Academy of Sciences) emphasized the importance of medical error and its deleterious impact on patient outcome.^{38,39} Clinical error rates have been reported from about 1% to 50%.^{38,40} The inaccuracy, with which skilled physicians perceive physiologic data while making clinical decisions, contributes substantially to these errors.⁴¹ Ill-defined terms or statements must also contribute.^{25,42,43} Strikingly, even when error rates in a carefully managed academic ICU were only 1%, every patient was subjected to a major threat to life or limb every other day, on average!⁴⁰ This error rate is lower than in most clinical environments. Very few clinical environments can claim to achieve a correct clinician performance rate of 99%.

The emerging problem of multimorbidity accentuates the complexity of the chronic care model in clinical practice in routine outpatient care. Forty-eight percent of elderly patients have at least three diseases, only 47% of arthritis patients have arthritis alone, and fewer than 20% of patients with coronary artery disease have it alone.⁴⁴ The number of variables, for example in COPD, is staggering.⁴⁵ Even when some trustworthy evidence of intervention benefit exists, benefits may easily get lost in the noise (unnecessary variation) of everyday medical practice, unless the treatment effects are large (an uncommon scenario). The more complex and “advanced” the medical problems, the higher the chances that medicine can become a nuisance to individuals and to public health, with harmful consequences, such as the extensive use of antiarrhythmic agents before the Cardiac Arrhythmia Suppression Trial.⁴⁶ How many established practices need to be abandoned remains a pertinent question.⁴⁷

ADAM SMITH’S INVISIBLE HAND AND CLINICIANS

Individual clinician decision-makers operate under the philosophical influence of Adam Smith’s “invisible hand.” The “invisible hand” theory assumes that even though players (here, clinicians) act on their own private intuition and preferences, somehow an “invisible hand” self-regulates the system to obtain maximal benefits for society at large. Many seem to believe, with quasireligious rather than scientific justification, that well-intended individual health-care practitioners will somehow eventually produce benefits for patients and society. However, the “invisible hand” does not appear to produce desired economic community benefit. “Rewards that depend on relative performance spawn collective action problems that can cause markets to fail.”⁴⁸ Similarly, the current approach has failed to produce desired community health benefits and many workers have called for an

enterprise-wide, or systems, approach.⁴⁹ Individualized clinician decision-making can clearly sometimes be good both for innovation and for individual patient care needs, but the common deviation of decisions from evidence-based care can also produce individual and community harm.^{25,38,39}

Modern medicine appears both inefficient, and costly, with health care poised to ruin many advanced societies. The United States pays more per person for health care than any Organization for Economic Co-operation and Development country but realizes less than the median life expectancy.⁵⁰⁻⁵² Current US health-care expenses appear unsustainable.⁵³ US health-care costs in 2011 are expected to be \$2.7 trillion, and it is estimated that at least one-third (\$900 billion) is spent for ineffective or unnecessary care.⁵⁰

UNNECESSARY VARIATION

As a result of the complexity of health-care delivery systems and the current model of individual clinician decision-making, unnecessary variation in clinical care is ever-present.^{31,54,55} Unnecessary variation is associated with unwanted and widespread error.^{38,39,56-61} The common abhorrence of “cookbook” (identical treatment of all patients) medicine can lead clinicians to conclude that clinical decision-making variability should be a desired goal. While this is intuitively attractive, this goal requires two assumptions.

First, that clinicians can properly tailor individual treatment to produce desirable outcomes at both the individual and the community-wide scale. Clinicians can, of course, tailor treatment to individual patients successfully, but they can also fail to deal correctly with the individualized needs of patients and thereby cause harm.^{38,39,59,62-73} Human decision-maker cognitive limitations^{65,73,74} make clinicians unable to easily predict who will respond to a specific intervention.^{69,70,75} Absent large signals, such as the clinical response of patients with pneumococcal pneumonia to penicillin in the 1940s, we can generally only draw conclusions about the balance between potential good and potential harm through examination of the results of credible systematic investigations.

Second, that nonuniformity fosters insight and innovation and is itself desirable. Bright and clever clinicians introduced many advances in medicine through careful observation. However, many modern medical questions involve small improvements with low (< 1.3) or tiny (< 1.05) ORs that escape the attention of most observers, absent rigorous systematic study.^{16,76} The success of stabilization of processes in nonmedical domains has provided a model for quality improvement in health care.⁷⁷⁻⁸¹ Stabilization of processes increases chances that small improvements may

be detected after system changes. Clinical process improvement (quality improvement) has adopted this approach.^{6,66,82-84} However, most clinical quality improvement studies rely on nonrandomized data and, thus, their ability to identify small effects remains tenuous.

HUMAN DECISION-MAKING LIMITATIONS

The health-care environment generates overwhelming amounts of information.^{18,25} Human decision-making ability is not well matched to the information commonly encountered by clinical decision-makers.^{65,73,74} The information overload that clinicians encounter probably contributes to both clinical decision-maker performance variability and high error rate.^{34,35,85-94}

The iterative interactions between patients and the clinical environment establish a transactional unit.⁹⁵ Patients express their illness or needs and the clinical caregivers respond, influencing the patients' expressions (phenotypes). These interactions between patients and caregivers are iterative, in contrast to the more simplified type of questions posed by the majority of clinical studies. This transactional unit is the proper unit of analysis for understanding outcomes and their determinants because it includes both the patient and the clinicians within the clinical care context.^{25,95} For example, the outcome of a baby born with cystic fibrosis depends on clinical context. If born in San Francisco, the infant will likely have a pulmonary disorder of young adult years. If the same infant is born in a third-world country rural area, it will likely have a lethal neonatal GI disease.

Central to the interactions (between patient and clinician) of this transactional unit, are human decision-making limitations. These limitations are linked to a human short-term memory limit of about 4 ± 1 constructs.^{35,96,97} Human decision-making limitations impede achievement of reproducibility in clinical experiments or care.²⁵ In addition, as biomedical information grows many health-care decisions may happen outside of the patient-clinician encounter. Patients are bombarded by direct-to-consumer advertising for drugs,⁹⁸ or personal "omics,"^{99,100} to name just two examples. Decision-making limitations of patients alone may be even worse than those of physicians in the traditional patient-clinician encounter, especially if the information being fed directly to patients is opportunistic, fragmented, conflicting, and conflicted.

LIMITATIONS OF THE CURRENT CLINICAL INVESTIGATION STRATEGY AT THE PATIENT-CLINICIAN ENCOUNTER SCALE

Decision-support tools, like guidelines and protocols, can support simple or complex care processes.¹⁰¹ Guidelines and protocols can increase compliance

with evidence-based interventions, support clinical decision-making, can reduce variation in clinical care,¹⁰² and influence clinician performance and patient outcome favorably.¹⁰³⁻¹⁰⁹ While not commonly formally studied, they likely reduce error as well.^{25,110} Simple protocols, like physician reminders for abnormal serum potassium measurements, are commonly used.¹¹¹ Complex protocols seem likely to have greater potential to aid clinicians and to reduce error.²⁵ However, the extent of publication and selective outcome reporting biases favoring guidelines and protocols is unknown, and may be substantial. Most of the research on decision-support tools has not been registered in advance. Their outcomes can be flexible and modified. Selective results reporting may be easy. Moreover, there is considerable evidence of infiltration of panels producing guidelines, and other decision-support tools, by influential experts with strong conflicts of interest.¹¹²⁻¹¹⁵ The majority of new guideline statements lack much trustworthy evidence and depend largely on expert judgment and influence.¹¹⁵

Perhaps an even more critical issue is how to identify the appropriate evidence for decision-support tools. Usually, clinical studies report the average effect of an intervention in heterogeneous subjects with different background characteristics and risks, and thus very different risk-benefit ratios and cost-benefit ratios.¹¹⁶ Stephen Gould¹¹⁷ has argued persuasively that many conclusions based on central tendency (usually the mean value) are likely to be incorrect. For example, the important signal for patients treated with insulin for stress hyperglycemia may reside in those subjects with very high values, those with very low values, or those with very rapid changes in blood glucose. These subgroups of subjects might be poorly reflected in the mean values of the experimental groups. Risk-based analyses using validated models have been proposed to better identify and model the impact of treatments on individual patients,^{118,119} but validated models may not always be available.

Treatment effect variability for different strata of patients may underlie some discrepancies commonly encountered between different studies of the same putative issue. When evidence cannot answer conclusively what should be done on average, it is even more difficult to draw any solid inferences on more limited strata, let alone realize the promises of individualized (personalized) medicine.^{120,121} Multiple recent studies of blood glucose management with IV insulin in critically ill patients were deemed of only fair quality in meta-analyses.¹²² In spite of uncertainties, these study results nevertheless led to an American College of Physicians guideline that could impact regulatory oversight of clinical care.¹²³ Trustworthy extrapolations of these data on specific strata and individuals are very difficult, if not impossible.

One could consider two major categories of decision-support tool use. The first is for clinical care, using the best evidence. The second is to stabilize the processes of clinical investigation, with or without available best evidence. Decision-support tools may be of limited help when the evidence is irreparably biased, or when clinical investigation has only provided information on misleading averages, and needed insights into stratified effects are absent or spurious. However, even when evidence is absent, decision-support tools can likely stabilize clinical investigation processes. Decision-support tools may enable better recognition of the effects of interventions, as proposed by Deming⁷⁹ and the clinical quality improvement movement. For example, adequately explicit (detailed) decision-support tools can produce replicable clinician decision-making for management of stress hyperglycemia between different clinical sites within different cultures.^{25,124-126} However, decision-support tools need to be tested on a large scale, and with multiple clinical care problems, to better evaluate their role in general health-care delivery.

Unfortunately, high-profile reviews of decision support do not assess the replicability of clinician decision-making.^{106,109} Neither reviews nor most individual studies examine clinician compliance with decision-support tool recommendations and thus do not examine the consistency of clinician decision-making in a given clinical context. Thus, it remains unclear how effective the reported decision-support tools are in stabilization of process, ie, leading different clinicians, when faced with the same clinical data, to the same decision. For example, one cannot ascertain the compliance of clinicians managing stress hyperglycemia with the insulin administration recommendations of the protocols in either the method of the compelling work of van den Berghe et al¹²⁷ or in that of the NICE-SUGAR study.¹²⁸ Without such compliance information, one cannot identify the method that was actually used. If clinicians rejected protocol recommendations in favor of their bedside clinician judgment, then their judgment became an unarticulated and unknown part of the experimental method.²⁵ Inconsistency in applying recommendations at the patient-clinician encounter scale may explain, in part, why the medical community is now no more informed about management of blood glucose than it was in 1999. After the seminal report by van den Berghe et al¹²⁷ of mortality reduction when clinicians aimed to bring blood glucose within an 80 to 110 mg/dL target with IV insulin in 2001, multiple studies have produced variable results.^{32,33,129-131} These concerns raise fundamental questions about our clinical experimental enterprise at the patient-clinician encounter scale.^{7,14,25}

Many clinical trials still suffer from low numbers of enrolled subjects.¹³² We should aim for much larger enrollments to answer clinically relevant questions, but this requires larger numbers of enrolling clinical sites. Good experimental design requires that all enrolling sites use identical methods. However, we lack widely available and adequately explicit decision-support tools that enable different clinicians in different sites to achieve the same experimental method. For example, we are currently unable to separate the effects of different insulin doses from those of blood glucose targets on outcomes of critically ill patients. We are also unable to generate dose-response curves for blood glucose targets or for insulin doses because of the low numbers of subjects enrolled in clinical trials. We might advance if we adopted an adequately explicit detailed Web-based computer protocol for management of blood glucose.^{25,125,126} Such a protocol used by 5,000 sites might enroll 200,000 subjects in 3 or 4 months. We could then execute a 5×5 factorial study design, with five insulin clamps (constant insulin infusions rates) and five blood glucose targets that would generate phase 3 human dose-response curves for both insulin doses and blood glucose targets. Results from such a study would address the independent effects of insulin doses and blood glucose targets on pertinent clinical outcomes. This proposed future investigative environment of megatrials with stabilized processes could more effectively address many comparative effectiveness research questions²⁰ and enable achievement of the Institute of Medicine goal of a learning health-care environment.¹⁰¹ These achievements will not likely be realized by widespread distribution of electronic health records alone without adequately explicit clinician decision support to stabilize the investigative methods. The noise and lack of consistency in applying interventions in such electronic health record observational data are likely to obscure real signals.

CONCLUSION AND FINAL COMMENTS

Rigorous, sound, efficient and replicable experimental methods are sought in all scientific domains because they are required for trustworthy study results. For most health-care delivery decisions, the most pertinent information is derived from the patient-clinician encounter scale. Obtaining a benefit at this scale requires human (clinician) behavior modification. Traditional strategies, including education, guidelines, and protocols, have led to only limited progress for many reasons, some of which we have discussed in the two preceding sections. Detailed computer decision-support tools that incorporate the best evidence may advance our research and care, but their value needs

to be documented in carefully conducted studies. Even in the absence of trustworthy evidence, adequately explicit decision-support tools may stabilize the clinician decision-making process and introduce uniformity among different clinicians while still allowing personalized care. Some serious rethinking and restructuring of our clinical research and care delivery systems will be necessary to assure the profession and the public that we continue to do more good than harm. Solutions to this challenge may require better clinical research methods that decrease the noise-to-signal ratio, permit fewer conflicts of interest, enable systematic use of trustworthy information, support careful implementation of this knowledge in decision-support tools, and provide appropriate oversight oversight. National and international programs to develop and evaluate decision-support tools that enable reproducible clinician behavior should be considered. We suspect such programs would fit nicely into current Comparative Effectiveness Research and Patient-Centered Outcomes Research Institute National Institutes of Health initiatives and may extend also to international efforts.

ACKNOWLEDGMENTS

Financial/nonfinancial disclosures: The authors have reported to *CHEST* that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Role of sponsors: The sponsors had no role in the design of the study, the collection and analysis of the data, or in the preparation of the manuscript.

Other contributions: We are indebted to Jason Christie, MD, for reviewing the manuscript.

REFERENCES

1. Beauchamp T, Childress J. *Principles of Biomedical Ethics*. 3 ed. New York, NY: Oxford University Press, Inc; 1989:470.
2. Sharpe V, Faden A. *Medical Harm: Historical, Conceptual, and Ethical Dimensions of Iatrogenic Illness*. New York, NY: Cambridge University Press; 1998.
3. Flexner S. *The Random House Dictionary of the English Language*. 2 ed. New York, NY: Random House Inc; 1987: 193-194.
4. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica*. 1979;47(2):263-291.
5. Kahneman D. *Thinking, Fast and Slow*. 1 ed. New York, NY: Farrar, Straus, and Giroux (Macmillan). Amazon website. <http://amazon.com>. Accessed October 25, 2011.
6. Berwick DM. Eleven worthy aims for clinical leadership of health system reform. *JAMA*. 1994;272(10):797-802.
7. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
8. Ioannidis JP. An epidemic of false claims. Competition and conflicts of interest distort too many medical findings. *Sci Am*. 2011;304(6):16.
9. Greene H, Roden DM, Katz RJ, et al. The Cardiac Arrhythmia Suppression Trial: first CAST...then CAST-II. *J Am Col Cardiol*. 1992;19:894-898.
10. Price HH. Hume's analysis of belief. Lecture 7. Belief 1959-61. Gifford Lectures website. <http://www.giffordlectures.org/>

- Browse.asp?PubID=TPBELI&Volume=0&Issue=0&ArticleID=8. Accessed July 2012.
11. Jasny BR, Chin G, Chong L, Vignieri S. Data replication & reproducibility. Again, and again, and again....introduction. *Science*. 2011;334(6060):1225.
12. Peng RD. Reproducible research in computational science. *Science*. 2011;334(6060):1226-1227.
13. Lehrer J. The truth wears off: is there something wrong with the scientific method? *New Yorker*. December 13, 2010: 52-57.
14. Ioannidis JPA, Khoury MJ. Improving validation practices in "omics" research. *Science*. 2011;334(6060):1230-1232.
15. Siontis GC, Ioannidis JP. Risk factors and interventions with statistically significant tiny effects. *Int J Epidemiol*. 2011;40(5):1292-1307.
16. Siontis KC, Evangelou E, Ioannidis JP. Magnitude of effects in clinical trials published in high-impact general medical journals. *Int J Epidemiol*. 2011;40(5):1280-1291.
17. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol*. 2011;64(10):1060-1069.
18. Committee on Patient Safety and Health Information Technology (Board on Health Care Services). *Health IT and Patient Safety: Building Safer Systems for Better Care*. Academies IoMotN, ed. Washington, DC: Institute of Medicine of the National Academies; 2012:197.
19. Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, ed. *Clinical Practice Guidelines We Can Trust*. Washington, DC: National Academies Press; 2011.
20. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med*. 2009;151(3):203-205.
21. von Bertalanffy L. *General System Theory*. New York, NY: George Braziller; 1968.
22. Miller J. *Living Systems*. New York, NY: McGraw-Hill Book Company; 1978.
23. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge, England: Cambridge; 2000:384.
24. Campbell D, Stanley J. *Experimental and Quasi-Experimental Designs for Research* (reprinted from Handbook of Research on Teaching, 1963). Boston, MA: Houghton Mifflin Co; 1966.
25. Morris A. The importance of protocol-directed patient management for research on lung-protective ventilation. In: Dreyfuss D, Saumon G, Hubamyr R, eds. *Ventilator-Induced Lung Injury*. New York, NY: Taylor & Francis Group; 2006: 537-610.
26. Lavizzo-Mourey L. It's time to connect what we know with what we do. In: United Health Foundation the American Public Health Association and Partnership for Prevention, ed. *America's Health Rankings*. Minnetonka, MN: United Health Foundation; 2008:7-8.
27. A call to action for individuals & their communities. United Health Foundation the American Public Health Association and Partnership for Prevention, ed. *America's Health Rankings*. Minnetonka, MN: United Health Foundation; 2008:107.
28. Cronin L, Cook DJ, Carlet J, et al. Corticosteroid treatment for sepsis: a critical appraisal and meta-analysis of the literature. *Crit Care Med*. 1995;23(8):1430-1439.
29. Lefering R, Neugebauer EA. Steroid controversy in sepsis and septic shock: a meta-analysis. *Crit Care Med*. 1995; 23(7):1294-1303.
30. Cook DJ, Reeve BK, Guyatt GH, et al. Stress ulcer prophylaxis in critically ill patients. Resolving discordant meta-analyses. *JAMA*. 1996;275(4):308-314.

31. Wennberg JE. Unwarranted variations in healthcare delivery: implications for academic medical centres. *BMJ*. 2002; 325(7370):961-964.
32. Singh JA, Hodges JS, Toscano JP, Asch SM. *Quality of care for gout in the US needs improvement*. *Arthritis Rheum*. 2007;57(5):822-829.
33. Wiener RS, Wiener DC, Larson RJ. Benefits and risks of tight glucose control in critically ill adults: a meta-analysis. *JAMA*. 2008;300(8):933-944.
34. Abramson NS, Wald KS, Grenvik ANA, Robinson D, Snyder JV. Adverse occurrences in intensive care units. *JAMA*. 1980;244(14):1582-1584.
35. Wu AW, Folkman S, McPhee SJ, Lo B. Do house officers learn from their mistakes? *JAMA*. 1991;265(16):2089-2094.
36. Reason J. *Human Error*. Cambridge, England: Cambridge University Press; 1990.
37. Reason J. Human error: models and management. *BMJ*. 2000;320(7237):768-770.
38. Kohn L, Corrigan J, Donaldson M, eds. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
39. Committee on Quality of Health Care in America, ed. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
40. Gopher D, Olin M, Badihi Y, et al. The nature and causes of human errors in a medical intensive care unit. In: *Proceedings from the Human Factors and Ergonomics Society Annual Meeting*; October 16-20, 1989; Denver, CO; 33: 956-960.
41. Evaluation of right heart catheterization in critically ill patients. Ontario Intensive Care Study Group. *Crit Care Med*. 1992;20(7):928-933.
42. Guidelines for the care of patients with hemodynamic instability associated with sepsis. Guidelines Committee; Society of Critical Care Medicine. *Crit Care Med*. 1992; 20(7):1057-1059.
43. Morris AH. Hemodynamic guidelines. *Crit Care Med*. 1993;21(7):1096.
44. Boyd C, Fortin M. Future of multimorbidity research: how should understanding of multimorbidity inform health system design? *Public Health Rev*. 2010;32(2):451-474.
45. Agustí A, Vestbo J. Current controversies and future perspectives in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2011;184(5):507-513.
46. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med*. 1989;321(6): 406-412.
47. Prasad V, Cifu A, Ioannidis JP. Reversals of established medical practices: evidence to abandon ship. *JAMA*. 2012; 307(1):37-38.
48. Frank RH. *The Darwin Economy: Liberty, Competition, and the Common Good*. Princeton, NJ: Princeton University Press; 2011:xvi,240.
49. Crowley WF Jr, Sherwood L, Salber P, et al. Clinical research in the United States at a crossroads: proposal for a novel public-private partnership to establish a national clinical research enterprise. *JAMA*. 2004;291(9):1120-1126.
50. Begley S. The best medicine. *Sci Am*. 2011;305(1):50-55.
51. Committee on Reviewing Evidence to Identify Highly Effective Clinical Services (Board on Health Care Services). *Knowing What Works in Health Care: A Roadmap for the Nation*. Academies IoMotN, ed. Washington, DC: Institute of Medicine of the National Academies; 2008:280.
52. Schraad-Tischler D. *Social Justice in the OECD—How Do the Member States Compare? Sustainable Governance Indicators 2011*. Gütersloh, Germany: Bertelsmann Foundation; 2011.
53. American College of Physicians. *How Can Our Nation Conserve and Distribute Health Care Resources Effectively and Efficiently?* Philadelphia, PA: American College of Physicians; 2011.
54. Wennberg JE, Gittelsohn A. Small area variation analysis in health care delivery. *Science*. 1973;142:1102-1108.
55. Morris AH. Iatrogenic illness: a call for decision support tools to reduce unnecessary variation. *Qual Saf Health Care*. 2004;13(1):80-81.
56. Leape LL, Berwick DM, Bates DW. What practices will most improve safety? Evidence-based medicine meets patient safety. *JAMA*. 2002;288(4):501-507.
57. Kozier E, Seto W, Verjee Z, et al. Prospective observational study on the incidence of medication errors during simulated resuscitation in a paediatric emergency department. *BMJ*. 2004;329(7478):1321.
58. Schiff GD, Klass D, Peterson J, Shah G, Bates DW. Linking laboratory and pharmacy: opportunities for reducing errors and improving care. *Arch Intern Med*. 2003;163(8): 893-900.
59. Runciman WB, Merry AF, Tito F. Error, blame, and the law in health care—an antipodean perspective. *Ann Intern Med*. 2003;138(12):974-979.
60. Lamb RM, Studdert DM, Bohmer RM, Berwick DM, Brennan TA. Hospital disclosure practices: results of a national survey. *Health Aff (Millwood)*. 2003;22(2):73-83.
61. Zhang J, Patel VL, Johnson TR. Medical error: is the solution medical or cognitive? *J Am Med Inform Assoc*. 2002; 9(suppl 6):S75-S77.
62. Silverman WA. Doing more good than harm. *Ann N Y Acad Sci*. 1993;703:5-11.
63. Warren K, Mosteller F, eds. *Doing More Good Than Harm: The Evaluation of Health Care Interventions*. New York, NY: The New York Academy of Sciences; 1993.
64. Horwitz RI, Singer BH, Makuch RW, Viscoli CM. Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation. *J Clin Epidemiol*. 1996; 49(4):395-400.
65. Redelmeier DA, Ferris LE, Tu JV, Hux JE, Schull MJ. Problems for clinical judgement: introducing cognitive psychology as one more basic science. *CMAJ*. 2001;164(3): 358-360.
66. Berwick DM. Disseminating innovations in health care. *JAMA*. 2003;289(15):1969-1975.
67. Barach P, Berwick DM. Patient safety and the reliability of health care systems. *Ann Intern Med*. 2003;138(12): 997-998.
68. Sox HC. Improving patient care. *Ann Intern Med*. 2003; 138(12):996.
69. Senn S. Individual response to treatment: is it a valid assumption? *BMJ*. 2004;329(7472):966-968.
70. Lake APJ. Every prescription is a clinical trial. *BMJ*. 2004; 329(7478):1346.
71. Corke CF, Stow PJ, Green DT, Agar JW, Henry MJ. How doctors discuss major interventions with high risk patients: an observational study. *BMJ*. 2005;330(7484):182.
72. Berwick DM. Improving patient care. My right knee. *Ann Intern Med*. 2005;142(2):121-125.
73. Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. *Ann Intern Med*. 2005; 142(2):115-120.
74. Kahneman D, Slovic P, Tversky A. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press; 1982.

75. MacLaren G, Combes A, Bartlett RH. Contemporary extracorporeal membrane oxygenation for adult respiratory failure: life support in the new era. *Intensive Care Med.* 2011;38(2):210-220.
76. Hulley S, Cummings S. *Designing Clinical Research.* Baltimore, MD: Williams and Wilkins; 1988.
77. Shewart W. *Economic Control of Quality of Manufactured Product.* New York, NY: D. Van Nostrand Co, Inc; 1931. Republished: Milwaukee, WI: American Society for Quality Control; 1980.
78. Deming W. *Quality, Productivity, and Competitive Position.* Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study; 1982.
79. Deming W. *Out of the Crisis.* Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study; 1986.
80. Imai M. *Kaizen—The Key to Japan's Competitive Success.* New York, NY: McGraw-Hill Publishing Company; 1986.
81. Walton M. *The Deming Management Method.* New York, NY: Putnam Publishing Group (Perigee Books); 1986.
82. James B, Horn S, Stephenson R. Management by fact: What is CPI and how is it used? In: Horn S, Hopkins D, eds. *Clinical Practice Improvement: A New Technology for Developing Cost-Effective Quality Health Care.* New York, NY: Faulker & Gray, Inc; 1994:39-54.
83. Horn S, Hopkins D, eds. *Clinical Practice Improvement: A New Technology for Developing Cost-Effective Quality Health Care.* New York, NY: Faulker & Gray, Inc; 1994.
84. James BC, Hammond ME. The challenge of variation in medical practice. *Arch Pathol Lab Med.* 2000;124(7):1001-1003.
85. Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge, England: Cambridge University Press; 1982:163-178.
86. Jennings D, Amabile T, Ross L. Informal covariation assessment: Data-based versus theory-based judgments. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge, England: Cambridge University Press; 1982:211-230.
87. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. *N Engl J Med.* 1976;295(24):1351-1355.
88. Morris AH, Chapman RH, Gardner RM. Frequency of technical problems encountered in the measurement of pulmonary artery wedge pressure. *Crit Care Med.* 1984;12(3):164-170.
89. Morris AH, Chapman RH, Gardner RM. Frequency of wedge pressure errors in the ICU. *Crit Care Med.* 1985;13(9):705-708.
90. Morris AH. Elimination of pulmonary wedge pressure errors commonly encountered in the ICU. *Cardiology.* 1985;30(10):941-943.
91. Iberti TJ, Fischer EP, Leibowitz AB, Panacek EA, Silverstein JH, Albertson TE; Pulmonary Artery Catheter Study Group. A multicenter study of physicians' knowledge of the pulmonary artery catheter. *JAMA.* 1990;264(22):2928-2932.
92. Iberti TJ, Daily EK, Leibowitz AB, Schecter CB, Fischer EP, Silverstein JH; The Pulmonary Artery Catheter Study Group. Assessment of critical care nurses' knowledge of the pulmonary artery catheter. *Crit Care Med.* 1994;22(10):1674-1678.
93. Leape LL. Error in medicine. *JAMA.* 1994;272(23):1851-1857.
94. Gnaegi A, Feihl F, Perret C. Intensive care physicians' insufficient knowledge of right-heart catheterization at the bedside: time to act? *Crit Care Med.* 1997;25(2):213-220.
95. Altman I, Rogoff B. World views in psychology: trait, interactional, organismic, and transactional perspectives. In: Stokols D, Altman I, eds. *Handbook of Environmental Psychology.* New York, NY: John Wiley & Sons; 1987:7-40.
96. Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev.* 1956;63(2):81-97.
97. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci.* 2001;24(1):87-114.
98. Howell JV. Direct to consumer advertisement: the world of the market place. *BMJ.* 2007;335(7622):683-684.
99. Ransohoff DF, Khoury MJ. Personal genomics: information can be harmful. *Eur J Clin Invest.* 2010;40(1):64-68.
100. Gulcher J, Stefansson K. Genetic risk information for common diseases may indeed be already useful for prevention and early detection. *Eur J Clin Invest.* 2010;40(1):56-63.
101. Young PL, Olsen L. Roundtable on Evidence-Based Medicine, Institute of Medicine. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, DC: The National Academies Press; 2010.
102. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet.* 1993;342(8883):1317-1322.
103. Safran C, Rind DM, Davis RB, et al. Effects of a knowledge-based electronic patient record in adherence to practice guidelines. *MD Comput.* 1996;13(1):55-63.
104. Grimm RH Jr, Shimoni K, Harlan WR Jr, Estes EH Jr. Evaluation of patient-care protocol use by various providers. *N Engl J Med.* 1975;292(10):507-511.
105. Wirtschafter DD, Scalise M, Henke C, Gams RA. Do information systems improve the quality of clinical research? Results of a randomized trial in a cooperative multi-institutional cancer group. *Comput Biomed Res.* 1981;14(1):78-90.
106. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med.* 1994;120(2):135-142.
107. East TD, Heermann LK, Bradshaw RL, et al. Efficacy of computerized decision support for mechanical ventilation: results of a prospective multi-center randomized trial. *Proc AMIA Symp.* 1999;251-255.
108. Mullett CJ, Evans RS, Christenson JC, Dean JM. Development and impact of a computerized pediatric anti-infective decision support program. *Pediatrics.* 2001;108(4):E75.
109. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA.* 2005;293(10):1223-1238.
110. Morris AH. Decision support and safety of clinical environments. *Qual Saf Health Care.* 2002;11(1):69-75.
111. Hoch I, Heymann AD, Kurman I, Valinsky LJ, Chodick G, Shalev V. Countrywide computer alerts to community physicians improve potassium testing in patients receiving diuretics. *J Am Med Inform Assoc.* 2003;10(6):541-546.
112. Jones DJ, Barkun AN, Lu Y, et al; International Consensus Upper Gastrointestinal Bleeding Conference Group. Conflicts of interest ethics: silencing expertise in the development of international clinical practice guidelines. *Ann Intern Med.* 2012;156(11):809-816. W-283.
113. Norris SL, Burda BU, Holmer HK, et al. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. *J Clin Epidemiol.* 2012;65(7):725-733.

114. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. *BMJ*. 2011;343:d5621.
115. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines [published correction appears in *JAMA*. 2009;301(15):1544]. *JAMA*. 2009;301(8):831-841.
116. Ioannidis JPA, Garber AM. Individualized cost-effectiveness analysis. *PLoS Med*. 2011;8(7):e1001058.
117. Gould SJ. *Full House: The Spread of Excellence From Plato to Darwin*. 1st Harvard University Press, ed. Cambridge, MA: Belknap Press of Harvard University Press; 2011.
118. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298(10):1209-1212.
119. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials*. 2010;11:85.
120. Ioannidis JP. Invited commentary-Genetic prediction for common diseases. *Arch Intern Med*. 2012;172(9):744-746.
121. Ioannidis JPA. Limits to forecasting in personalized medicine: an overview. *Int J Forecast*. 2009;25(4):773-783.
122. Kansagara D, Fu R, Freeman M, Wolf F, Helfand M. Intensive insulin therapy in hospitalized patients: a systematic review. *Ann Intern Med*. 2011;154(4):268-282.
123. Qaseem A, Humphrey LL, Chou R, Snow V, Shekelle P. Use of intensive insulin therapy for the management of glycemic control in hospitalized patients: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2011;154(4):260-267.
124. Thompson BT, Orme JF, Zheng H, et al; Reengineering Critical Care Clinical Research Investigators. Multicenter validation of a computer-based clinical decision support tool for glucose control in adult and pediatric intensive care units. *J Diabetes Sci Tech*. 2008;2(3):357-368.
125. Morris AH, Orme J, Rocha BH, et al; Reengineering Critical Care Clinical Research Investigators. An electronic protocol for translation of research results to clinical practice: a preliminary report. *J Diabetes Sci Tech*. 2008;2(5):802-808.
126. Morris AH, Orme J Jr, Truweit JD, et al. A replicable method for blood glucose control in critically ill patients. *Crit Care Med*. 2008;36(6):1787-1795. Epub 2008/06/04.
127. van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in critically ill patients. *N Engl J Med*. 2001;345(19):1359-1367.
128. The NICE-SUGAR Study Investigators, Finfer S, Chittock DR, Su SY, et al. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009;360(13):1283-1297.
129. Srinivasan V. Blood glucose variability in critical illness: Is it time to cast a wider net? *Pediatr Crit Care Med*. 2008;9(4):441-442.
130. Inzucchi SE, Siegel MD. Glucose control in the ICU—how tight is too tight? *N Engl J Med*. 2009;360(13):1346-1349.
131. Egi M, Bellomo R, Reade MC. Is reducing variability of blood glucose the real but hidden target of intensive insulin therapy? *Crit Care*. 2009;13(2):302.
132. Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet*. 2005;365(9465):1159-1162.