

Health Systems, Heuristics, and Comparative Effectiveness Research

Robert H. Brook, MD, ScD^{1,2}

¹RAND Corporation, Santa Monica, CA, USA; ²David Geffen UCLA School of Medicine, Los Angeles, CA, USA.

J Gen Intern Med 28(2):172–3

DOI: 10.1007/s11606-012-2242-y

© Society of General Internal Medicine 2012

All health professionals strive to do no harm and help their patients achieve their desired health status. To pursue these aims, physicians rely extensively on data, much of which in the near future may come from comparative effectiveness studies. About \$500 million a year is being spent on such studies in the US; multiple papers will appear each year in leading medical journals, concluding that therapy A can, or cannot, be distinguished from therapy B.

Will physicians be able to use such information to arrive at better decisions for their patients? Two recent books—*The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*¹ and *Thinking, Fast and Slow*²—demonstrate how difficult it will be to use comparative effectiveness data appropriately.

In the *Improvement Guide*, Gerald Langley and his co-authors discuss how to distinguish between a process—for example, performing coronary artery bypass surgery—that is unstable and one that is stable. If a process such as performing an operation is unstable (e.g., results are worse when the operation is performed in the morning or on a Friday), then the reason for this unstable pattern can be isolated and hopefully altered (e.g., the morning operating room team is missing a vital piece of equipment, which is replaced). But if the process is stable—for example, the results do not vary by the day of week or time or surgeon—but the results are still not acceptable, then a process that affects all patients undergoing a coronary artery bypass operation, such as how a patient is managed in a recovery room, must be changed to produce a better outcome (i.e., lower mortality).

Performing isolated coronary artery bypass surgery was, in the past, an unstable process in which outcomes varied depending on the surgeon or the hospital in which the surgery took place. However, this process has recently become stable. States such as California³ and New York⁴ produce public reports that link mortality from this procedure with individual surgeons or specific hospitals.

Over the last decade, mortality from the procedure has declined remarkably to less than 2%. It is now almost impossible to identify a surgeon or hospital in either state that is better or worse than other surgeons or hospitals. The process is stable, and future reductions in mortality will require finding a way to improve care in all hospitals instead of focusing on a particular hospital.

The second book—*Thinking, Fast and Slow*—describes a different challenge in using research results, especially comparative effectiveness research results that show small differences between the alternative therapies. Author Daniel Kahneman describes two modes of thinking. “System 1” thinking corresponds to fast, intuitive, emotional, and almost automatic decisions—flawed but necessary thinking. “System 2” thinking is slower and requires more intellectual effort. These two systems interact to produce human judgment.

Kahneman provides multiple examples of flawed thinking that have implications for how medicine is practiced. For instance, a person’s hands are immersed in painfully cold water for 60 s, then the hands are withdrawn. The experiment is repeated, but this time the hands are immersed for 60 s in water just as cold as in experiment 1, then for an additional 30 s in water cold enough to cause pain but less pain than was caused in the first 60 s. Then the person is asked whether he or she would “prefer” the first immersion (60 s of pain) or the second (90 s of pain, consisting of 60 s of *extremely* cold water plus 30 s of *very* cold water). The individual invariably picks 90 s. Thus, the level of pain at the end of a procedure trumps the duration of pain or total amount of pain. How can clinical medicine be designed to take advantage of this less-than-rational thinking?

In a second example, Kahneman shows how flawed thinking can make a system worse. A captain is charged with supervising the training of fighter pilots. One pilot completes the training routine perfectly and is congratulated by the captain. A second pilot does a miserable job and nearly crashes his plane. The captain chastises this pilot. In the next round of training, the first pilot performs less well, and the second pilot does better. The captain thus concludes that punishment is more effective than reward for changing behavior.

However, what did the captain actually observe? The first performance of the first pilot was due to both skill and luck.

Published online October 12, 2012

Similarly, the performance of the pilot who did poorly on the first pass was due to slightly less skill and worse luck (Kahneman uses the word luck, but luck could be a word that is just a placeholder for a phenomenon about which the captain was unaware—e.g., a solar flare that affected the plane's instruments). In his second pass, the pilot was luckier. In essence, luck is a fundamental force shaping the outcomes of processes all around us. When luck is not considered in how we choose to reward or punish, we communicate inappropriate messages to people with whom we work and make system performance worse. For instance, consider the interactions between attending physicians and residents in morning rounds. Does the teaching style resemble that of the captain, with a resident who had an unlucky night being punished?

In the original comparative effectiveness study conducted 40 years ago that compared coronary artery bypass surgery to medical therapy, risk-adjusted mortality rates for surgical patients varied 20 fold across the study sites.⁵ In Langley's terms, the process was unstable because the hospital in which the patient received the operation mattered. If every site had had a risk-adjusted mortality rate that resembled the best site, would physicians' recommendations about surgery versus medical care have been different?

In addition, since the study was conducted, complication rates for both surgery and medical therapy have fallen dramatically; the systems for performing surgery and providing medical therapy have improved. Is it appropriate to apply today the results of a comparative effectiveness study conducted 40 years ago when both therapies were very different and at least the surgical intervention was unstable?

Finally, Kahneman demonstrates that context and emotion often dominate rational decision-making. For instance, an individual is given the option of "choosing" surgery versus angioplasty and the former is described as cracking open the chest, the latter is described as inserting a small catheter in a vessel. Then after hearing these descriptions, can the patient actually use facts about outcomes when he makes a decision?

Considering both unstable-stable system issues and the probability of flawed thinking on the part of patients, should physicians believe they can use data to help patients choose between therapies that science says produce slightly different results? Most comparative effectiveness studies are conducted in unstable systems, where there are multiple sources of outcomes variation. It is impossible to interpret the results of these studies appropriately without knowing the sources of those variations and their effects on outcomes.

Sometimes, of course, a few studies could produce such dramatic results that the above considerations would be mostly irrelevant. But this would be rare.

What if at the very least comparative effectiveness research were performed only in systems that were stable? For example, how can a physician correctly use data from a study that compares medical therapy and carotid endarterectomy when the operation is performed differently by different surgeons, with different complication and mortality rates, and the physician does not know the complication rate of the surgeon to whom he will refer the patient? How can one type of anticoagulant be compared to another unless we understand a facility's commitment to manage a patient's anticoagulant status?

Until we can get our complex health care system working reliably, it's hard to imagine adding another layer in which, in the name of evidence-based medicine, one therapy is considered marginally better than another.

The extraordinary complexity of our world stems not only from the organizations and processes that engulf us, but also from the way physicians and patients interact with them. In the case of health care, the system produces an unreliable product. Yet researchers are trying desperately to prove in a bullet-proof manner that therapies provided in different health care environments are similar and that small differences between therapies are real. This is crazy making.

Getting patients to make rational choices between slightly different therapies may be impossible. Establishing that slightly different therapies are different may be next to impossible. But understanding our health system well enough to control it may be possible and can save many lives.

Corresponding Author: Robert H. Brook, MD, ScD; RAND Corporation, 1776 Main Street, Santa Monica, CA 90401, USA (e-mail: robert_brook@rand.org).

REFERENCES

1. **Langley GJ, Moen R, Nolan KM, Nolan TW, Norman CL, Provost LP.** *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance.* San Francisco, California, USA: Jossey-Bass Publishers; 2009.
2. **Kahneman D.** *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux; 2011.
3. *The California Report on Coronary Artery Bypass Graft Surgery: 2009 Hospital Data.* California CABG Outcomes Reporting Program; Office of Statewide Health Planning and Development, April 2012.
4. **Hannan EL, Cozzens K, King SB, Walford G, Shah NR.** The New York State cardiac registries: History, contributions, limitations, and lessons for future efforts to assess and publicly report healthcare outcomes. *J Am Coll Cardiol.* 2012;59(25):2309–16. doi:10.1016/j.jacc.2011.12.051.
5. **Kennedy JW, Kaiser GC, Fisher LD, Fritz JK, Myers W, Mudd JG, Ryan TJ.** Clinical and angiographic predictors of operative mortality from the collaborative study in coronary artery surgery (CASS). *Circulation.* 1981;63:793–802. doi:10.1161/01.CIR.63.4.793.