

Can We Trust Best Practices? Six Cognitive Challenges of Evidence-Based Approaches

Devorah E. Klein, Marimo Consulting, LLC, Medford, Massachusetts,
David D. Woods, The Ohio State University, Columbus,
Gary Klein, MacroCognition, LLC, Washington, D.C., and
Shawna J. Perry, University of Florida, Gainesville

There is a growing popularity of data-driven best practices in a variety of fields. Although we applaud the impulse to replace anecdotes with evidence, it is important to appreciate some of the cognitive constraints on promulgating best practices to be used by practitioners. We use the evidence-based medicine (EBM) framework that has become popular in health care to raise questions about whether the approach is consistent with how people actually make decisions to manage patient safety. We examine six potential disconnects and suggest ways to strengthen best practices strategies.

Keywords: quality, safety, evidence-based medicine, decision making, expertise, training

INTRODUCTION

The concept of “best practices,” informed by empirical evaluation, has enormous potential for guiding practitioners in many different fields toward treatments that have been shown to work and away from folk remedies that have little to offer besides tradition. Data are more important than intuitions, or at least they should be. However, it is simplistic to conclude that data and research are sufficient to improve practice. Practitioners need to make a variety of decisions about how to interpret and apply the evidence. This article explores some of the cognitive challenges of establishing and applying best practices—cognitive challenges of relying

on data for treatment recommendations. Our goal is to explore the limitations of data-driven decision making from a cognitive engineering perspective and make suggestions to overcome these limitations.

To illustrate some of the difficulties with a best-practices regimen, we use the example of evidence-based medicine (EBM) in health care. EBM seeks to establish a set of best practices for clinicians to use based primarily on the results of scientifically rigorous research (Gray, 1996; Roberts & Yeager, 2004). The EBM approach to clinical practice is to identify a treatment of interest; conduct carefully controlled studies, ideally using a double-blind paradigm; determine the effectiveness of the treatment; and disseminate the results in the form of best practices in the form of rules, such as “If X, then do Y.” These rules can be readily applied and also used to evaluate compliance. EBM offers a way to get rid of expensive treatments that do not make a difference to patients. This EBM script is an antidote to anecdotal medicine and a counter to the limits of the experience of physicians.

The concept of EBM has some ambiguity, as proponents have been progressively backing away from their original pronouncements about how it was going to revolutionize medicine. Further, EBM has been adapted for financial purposes in order to deny payments for noncompliance. We recognize these kinds of issues, but they are peripheral to the goals of this article. We are interested in the basic strategy of using data to improve the quality of treatment. EBM, including guidelines and best practices, are specific manifestations of a more general effort to apply modernist, rationalist approaches to improve treatments in fields such as health care (Wears & Hunte, 2014).

Address correspondence to Devorah E. Klein, PhD, Senior Scientist at Marimo Consulting, LLC, 44 Capen St., Medford, MA 02155, Devorah@marimoconsulting.com.

Journal of Cognitive Engineering and Decision Making
2016, Volume 10, Number 3, September 2016, pp. 244–254
DOI: 10.1177/1555343416637520
Copyright © 2016, Human Factors and Ergonomics Society.

Cognitive engineering insights and methods can be useful for all of these efforts. We are using EBM in medicine as an exemplar because that is where the best-practices program has been pursued most extensively.

This article describes some of the challenging decisions practitioners face when they try to use EBM for medical decision making in clinical work. Timmermans and Berg (2003) have discussed the sociocultural context of EBM. Our interest is in the cognitive challenges that confront EBM in the context of actual practice and the complexity of patients and diseases. An additional goal is to consider ways to complement or extend EBM to help health care professionals make better care decisions for and with patients.

SIX COGNITIVE CHALLENGES OF EBM

For physicians who work in complex, dynamic environments with high time pressure, uncertainty, and risk, EBM may appear a salvation. However, these attributes of complex settings may make it more difficult to employ the principles of EBM. Based on the results of several decades of study on how people make decisions under time pressure, uncertainty, and complexity, as well as extensive work on patient safety (e.g., Cook, Render, & Woods, 2000; Perry & Wears, 2011), we have identified six cognitive challenges facing clinicians who want to apply EBM in their practice.

1. Characterizing problems
2. Gauging confidence in the evidence
3. Deciding what to do when the generally accepted best practices conflict with professional expertise
4. Applying simple rules to complex situations
5. Revising treatment plans that do not seem to be working
6. Considering remedies that are not best practices

In all six challenges, a rigid adherence to EBM can run counter to clinical practitioners' cognitive and conceptual strengths as they face uncertain medical situations. In its extreme forms, EBM suggests a conflict between the use of evidence and the use of experience. We believe that little is gained by postulating such a conflict, and we advocate a blend of evidence with hard-won expertise and informed intuitions.

Challenge 1: Characterizing problems

The notion that we can apply a rule—"If this condition arises, apply that treatment"—ignores the ability to judge whether the condition has arisen. Much, if not most, of the challenge is in understanding what is anomalous and the nature of the underlying problems (Klein, Pliske, Crandall, & Woods, 2005). Once a problem is accurately identified, selection of a treatment can be straightforward. Best-practices approaches, such as EBM, tackle the easier part of the equation—how to address each problem—and do not give adequate coverage to the detection and identification of the problem, the variability within a problem category, and the nature of other interacting problems, conditions, and treatments. As a result, they undervalue the diagnostic expertise that is essential for effective health care. For example, the following vital signs—blood pressure = 90/60, heart rate = 130, respiratory rate = 30, oxygen saturation = 88%—open an immediate list of situations of concern (sepsis, congestive heart failure, emphysema flare) that are sorted out by expertise as the physician takes into account what is the context and who is the patient, a newborn baby or an 85-year-old woman with diabetes and a cough? Clinicians use heuristics (useful tactics that gain responsiveness and robustness but fall short of optimality) and pattern recognition (requiring diverse experiences) to cope with complexity. Best practices rely on categorizations as a form of organizing knowledge. These generalizations are simultaneously valuable and limited when a practitioner is confronted with variety, time course, and interactions across disease conditions. Heuristics, pattern recognition, and other forms of expertise lose their vitality, adaptiveness, and context sensitivity when reduced to rules.

Challenge 2: Gauging Confidence in the Evidence

Many physicians would like to access a database of best practices to identify a preferred treatment regimen for a patient, but too often the choice of treatment depends on judging the quality and relevance of the evidence, judgments that often depend on experience and are made under uncertainty and time pressure.

Evidence by itself can mislead us. Witness the frequent revision of claims that had empirical

support but were modified or overturned by later evidence in fields such as cancer treatment (Kaiser, 2015). In “garden path” problems, early-arriving pieces of evidence points to one hypothesis, but later, more subtle patterns of evidence reveal a different problem is actually present (Woods & Hollnagel, 2006).

Even if we ignore the problem of studies that cannot be replicated, we also have to worry about negative findings that are used to dismiss treatments that do have value. The data collection may have been flawed because of variables that were not well understood at the time. A well-known example is the set of best practices for treating peptic ulcer disease, thought to arise from excess stomach acids, likely caused by stress. When Barry Marshall explored the hypothesis that ulcers were actually caused by *Helicobacter pylori*, he conducted a critical experiment searching for *H. pylori* in the gut of ulcers victims. But the search failed. None of the patients showed any sign of *H. pylori*. That should have scuttled the hypothesis, but the problem was that the lab technicians were discarding the cultures after only a few days, the standard procedures they used for strep infections. Once the lab technicians accidentally gave the cultures more time to grow, the link to *H. pylori* was demonstrated definitively (Marshall, 2005). And the best practice shifted from surgery to antibiotics.

Some of the gold-standard sources of EBM have proved to be misleading—the Framingham Heart Study was performed focusing only on White males. However, many of the key symptoms for differentiating a heart attack from other diseases (i.e., “feels like an elephant sitting on my chest”) are found in only 5% of women. After one emergency physician realized the consequences of these limitations, she explained her horrified reaction: “I think about all the women I sent home to their deaths because I was following the algorithm for best practices for chest pain” (personal communication, April 6, 2012).

Excessive faith in evidence by itself can also literally blind people (Chatterjee, 2015) and lead to fixation on one view, which impairs the search for and utilization of new, emerging, alternative indicators. The 1981 Nobel Prize in Medicine, given to Torsten Wiesel and David Hubel,

focused on their finding of a critical period of neuroplasticity for vision. On the basis of this research, clinicians stopped performing corrective surgery for children born with cataracts if the intervention occurred after the age of 8. However, a researcher in India decided to challenge this dogma. When children’s cataracts were removed and artificial lenses inserted, vision was at least partially restored. The original Nobel Prize research, performed on primates, had led the medical community astray.

Examples like these show that the implications of evidence are not clear-cut, and the search for evidence is ongoing. Researchers may be unaware of the variables that are obscuring the relationships they are studying. As a result, the medical community has to be continually prepared to revise its faith in evidence that may seem solid today. Increasingly, researchers are appreciating the value of being uncertain (Timmermans & Berg, 2003). This is a critical finding in cognitive engineering—performance depends on how well people, teams, and organizations revise assessments as new evidence comes in (Woods & Hollnagel, 2006). A clear-cut example occurred in the run-up to the *Columbia* space shuttle disaster as managers discounted new evidence about new kinds of safety risks (Woods, 2005). EBM appears to offer a reassuring set of stable best practices, but in reality, we find that yesterday’s best practices become questioned and eventually revised or even discredited. Although once established, these practices may take a long time to be modified or discarded even when new evidence builds up.

Challenge 3: Deciding What to Do When the Generally Accepted Best Practices Conflict With Professional Expertise

This conflict is an issue because the professional judgment of physicians is credible, and their expertise and intuitions need to be taken into account. Physicians build up expertise and acquire pattern repertoires over years of experience. When they have ample opportunities for feedback about their judgments, their intuitions—their use of experience to make pattern-based judgments—are valuable (Klein, 1998). Kahneman and Klein (2009) assert that

intuitions are useful under two conditions: a reasonably stable environment and an opportunity for people to learn from feedback. For example, the stock market does not constitute a reasonably stable environment, and we are highly skeptical about claims of expertise or intuition in selecting stocks for investment. Another example is organizational decision making. Most people who work on the administrative side of organizations fail to get frequent, consistent, or accurate feedback and thus fail to develop expertise and credible intuition.

In contrast, medicine satisfies both of the conditions for credible intuitions. It is a reasonably stable environment, and physicians do get some feedback. Intuition here is not a random or mystical process but simply the use of pattern matching that is based on experience in a reasonably stable environment.

Admittedly, physicians do not achieve the levels of expertise found in chess grandmasters. Chase and Simon (1973) described how chess grandmasters accumulate tens of thousands of patterns that enabled them to rapidly size up situations. Chess is a highly stable environment—the positions of the pieces are unambiguous. And chess players receive clear feedback about the quality of their decisions. They know that they have won or lost a game and can go over the moves to determine where they made mistakes. Physicians do not receive the same level of feedback on their decisions. When they refer patients to specialists, they may not be informed about the results. Worse, feedback is abundant on common conditions but limited on rare conditions, such as early rabies. And worse yet, not all feedback is equal. Losses loom larger than successes. Ghafarzadegan, Epstein, and Martin (2013) have shown that a false negative (e.g., failing to do a needed C-section) will have a greater impact than a false positive (doing an unnecessary C-section). That is why physicians are sometimes cautioned not to trust their intuition. Indeed, this warning is part of the rationale for EBM.

We are not advocating for physicians to uncritically trust their intuitions. Rather, we see the importance of taking judgments and intuitions into account when making decisions, which will occasionally generate a conflict between best practices and professional judgment. Even though

medicine is not as stable and well structured a domain as chess, and does not permit the same quality of feedback, it is not a zero-reliability domain, like stock selection. Experience lets skilled physicians identify reasonable courses of action and reasonable treatments upon diagnosing a condition. These intuitions can sometimes be misleading, which is why they need to be checked by deliberately scrutinizing the conditions and likely consequences of different actions. Still, in complex situations, heuristics and pattern recognition are essential and cannot be replaced by sets of rules (DeAnda & Gaba, 1991; Wears & Schubert, 2015).

Challenge 4: Applying Simple Rules to Complex Situations

Problem solving is applied to both well-ordered and complex situations. Well-ordered situations are highly stable and are easily captured as a set of procedures. Insertion of a central line is a well-ordered situation in which a checklist approach has markedly reduced line infections (Pronovost et al., 2006). A central line is inserted into a large vein to deliver blood, medications, fluids, or nutrients for an extended period of time. In situations like this, there are fairly unambiguous tasks to be accomplished and clear criteria for success.

Complex situations, in contrast, contain many variables that must be taken into account. These variables all interact and vary over time as the patient's status changes. Further, a patient may be suffering from several different problems, making the patient's current status difficult to assess. A patient may have both asthma and diabetes. Severe asthma requires the use of steroids, but the steroids will drive up blood glucose. The physician cannot rely on either an asthma protocol or a diabetes protocol alone but will have to take both conditions into account, performing trade-offs that also reflect the characteristics and lifestyle of the individual patient. Much of health care involves wicked problems (Rittel & Weber, 1973) that do not have an unambiguously correct solution, a right treatment—but even so, some solutions are better than others.

Rules and evidence are about populations, but physicians have to treat individual patients. The evidence may focus on average patient

response data, downplaying the distribution and ignoring considerable individual variations. A given drug may be ineffective when averaged over an entire sample but may work at one of the extremes of this sample.

Physicians have to connect the general or categorical guidance of best practices to fit the individual patient in front of them. And that is difficult because the research base consists of studies that vary one thing at a time, so it is hard to adapt the findings to all of the factors present in an individual patient. The research base cannot vary all of the relevant factors without quickly running into combinatorial explosions and confounded variables.

Global rules do not necessarily apply blindly to specific patients. Patients, diseases, and comorbidities are more variable than the categories of best practices. As a result, no set of rules by themselves can completely specify how to treat individual patients under complex conditions. This finding holds regardless of whether sets of rules are organized as best practices, written as procedures, or mechanized in computers; rules are resources for considered action (Suchman, 1987; Woods, Roth, & Bennett, 1990). Skilled physicians certainly use statistically based generalization, but they also use other processes and inputs for situated cognition and action (Suchman, 1987).

One of the primary attractions of best practices is simplicity—the potential to dial in a course of treatment once a diagnosis is made. However, simplicity is also a limitation when confronting the complexity of individual patients.

Challenge 5: Revising Treatment Plans That Do Not Seem to Be Working

Health care providers often must rapidly determine when a treatment plan needs to be adapted. Plans have to be adapted for several reasons: They may not be working, complex situations are continually changing, and patient status may fluctuate. Plans are often created for these “wicked problems,” involving ill-defined goals that need to be revised. Medicine is practiced with patients who improve or deteriorate or develop additional symptoms and problems over time.

However, EBM is punctate; given these data, here are the recommendations. EBM is not well suited for plan adaptation. Practicing physicians who want to adhere to EBM have to wrestle with several challenges: the need to pick up early signs that a best practice is not producing the intended effects, early signs of expectancies that have been violated. They have to determine when and how to gather evidence to test these concerns and when to revise or withdraw a best practice.

Plan revision places great demands on expertise in understanding the treatment regimen and the individual patient so that revisions can be made quickly and effectively. Physicians need skill in revising a treatment plan without being too impatient to give it a chance to work or waiting so long that the patient’s chances of recovery are diminished. Rudolph, Morrison, and Carroll (2009) has examined the dynamic of hanging onto a plan for a while but being prepared to drop it when required. Research on dynamic and adaptive decision making can inform medical practice and be extended to enhance patient safety (Amalberti, 2013; Brehmer, 1987, 1992; Kylesten, 2013).

Woods (1994) and Woods and Hollnagel (2006) model anomaly response and plan revision based on studies in abnormal events in aviation, nuclear power emergencies, and space shuttle missions. As anomalies occur, judgments about when and how to revise a plan are difficult. Gaba, Maxwell, and DeAnda (1987) documented this need for adaptation in their study of anesthesia crises. Klein (2007a, 2007b) has found that revising a plan can be more difficult than initiating it because once a plan is under way, it can be difficult to disentangle the condition from the treatment. The patient may be getting worse or might be having a bad reaction to the medication even as the underlying disease is being brought under control. Plan revision can consist of changing a schedule to a more serious modification of adding or deleting tasks, to an even more serious step of replacing one strategy with another in order to achieve the goals, and in the most extreme cases, to revise or replace the goals themselves. Applying EBM assumes the best practice has the desired effect and fails to speak to the cognitive and teamwork issues in plan revision.

Challenge 6: Considering Remedies That Are Not Best Practices

Best-practice guidelines may not be available that cover all of the situations physicians face or may not address all aspects of a patient's case, yet physicians still have to make treatment decisions. How should a practitioner use other evidence not derived from controlled studies that seems relevant or fill gaps for the patients they encounter? We appreciate that not all EBM advocates insist on using evidence only when derived from double-blinded studies and that in many circumstances, it will be impossible to set up this kind of design. This raises the question of the role of less rigorously documented practices. Should physicians not apply remedies unless they are part of prescribed/sanctioned "best practices" even when practice turns out to be incomplete or insufficient for the patient who is suffering now? How should physicians act now when research trials to expand best-practices knowledge may not be completed for years? An all-or-none stance on rigor leaves a gap between the development of rigorous knowledge and the need to act now when knowledge is incomplete. And what to do about meta-analyses that throw out all information from studies that fail to meet the highest levels of rigor even though the study may provide some provisional information? Studies may fail to generate significant results because of variability in the data, but that variability may reflect subpopulations that have different reactions to treatments. Thus, meta-analyses can mask different reactions among various subpopulations.

These examples identify another judgment that EBM downplays but has been studied in cognitive engineering—the sufficient-rigor judgment (Zelik, Patterson, & Woods, 2010). Although one would like to make decisions based only on the most rigorous evidence, in the real world, there are always constraints from limited resources and time pressure. This means in practice there is a judgment about what level of rigor is sufficient at that point in time given the uncertainty and the risks of acting too late or too early. There are ways to assess and assist the sufficient-rigor judgment as decision makers struggle under pressure.

A related challenge is the role of practitioners in providing evidence. Does the evidence base

always fall within the provenance of medical researchers, or will data from other sources be accepted? Unexpected twists in a case, unexpected reactions, coincidences, all have contributed new hypotheses and ideas. Spectacular progress in surgery over the past century—joint replacement, cardiac valve replacement, reconstructions, and so forth—have been achieved without initial randomized control trials (O'Sullivan, 2010).

Finally, it seems foolish to ignore "rock stars"—individual practitioners who achieve much higher success rates than others. Staszewski (2004) provided an example of a rock star in the domain of mine detection. Conventional mine detectors proved ineffective in the face of a new generation of mines that relied on plastic parts instead of metal, so the Army invested \$38 million over 9 years to develop the next generation of handheld mine detector. Unfortunately, tests showed that the new version was not any better than the previous version; both achieved only 10% to 20% accuracy. However, there were a few specialists, one in particular, who were able to achieve success rates greater than 90% with the new equipment. By studying their strategies, it was possible to develop a training course that allowed soldiers to jump from 20% accuracy to 90% accuracy. When there are only a few experts, or just a single expert, statistical evaluation will be impossible. However, these standouts, sometimes referred to as positive deviants (Pascale & Sternin, 2010), offer important lessons. Of course, there is a risk of drawing the wrong lessons from standouts. For example, one risk of naive copying of successful cases involves undersampling failure (Denrell, 2003; Denrell & Fang, 2010). Organizations may try to learn from other, successful organizations and conclude that bold, risky decisions are essential. However, this conclusion misses the cases of organizations that failed because they took unfortunate risks; the failures are no longer in business and unavailable for study. The Staszewski example was not naive copying of standouts but was a detailed study to understand what made the standouts successful. In contrast, "best practice" as a paradigm in health care easily slides into naive copying. The above example from Staszewski also illustrates another aspect of experience and expertise.

Expertise is not expressible in the form of rule sets and so can be easily dismissed as mere intuition and unreliable. But forms of expertise, such as the recognition of patterns of relationships, can be identified, modeled, and supported (Hoffman et al., 2014).

DISCUSSION

Best-practices approaches oversimplify the cognitive challenges of putting general guidance into action when confronting specific complex situations under pressures. The world presents complexities and uncertainties that cannot all be overcome with sets of best practices. The variability of diseases and patients and the interactions across patient conditions spill over the category boundaries of best-practice guidance. Things do not always go as intended or planned, as harried clinicians treat patients with chronic conditions and comorbidities. The knowledge of what is “best” changes, and the sources of discovery of new knowledge are highly diverse and opportunistic.

In this context, EBM seems to overemphasize the limitations of heuristics, intuition, and expertise without appreciating their strengths. As a result, EBM advocates offer programs that seek to substitute for experience rather than augment it. A strict reliance on sanctioned evidence runs the risk of diminishing the experience and skills of practitioners rather than strengthening and calibrating those experiences, skills, and expertise (Hoffman et al., 2014). We are concerned that practitioners in a variety of disciplines may have trouble gaining expertise if they just mechanically apply prescribed rules, because this has occurred in other areas, such as aviation, where there has been deskilling/loss of expertise of pilots’ ability to handle non-normal and abnormal flight situations as overreliance on automation has grown (Abbott, McKenney, & Railsback, 2013).

Evidence-based best-practices programs have become a valuable antidote to anecdotal practice. Evidence, however, does not speak for itself. It needs to be interpreted, revised, and tailored to specific contexts and conditions, all of which takes expertise. In health care, the best-practices approach of EBM works best in well-ordered situations—for example, for tasks such

as inserting a central line, where the focus is on the task, not on the patient, and context is largely irrelevant. Problems arise when these successes encourage the medical community to establish best practices for complex situations that depend on interpreting subtle cues about specific patients, for instance, whether or not a patient is in good enough condition to be transitioned safely from the recovery room or intensive care unit to other levels of care (Cook, 2006).

EBM works smoothly when the evidence is clear and directly applicable to a patient. What about the more challenging situations in which physicians have to make decisions without good evidence? What about a patient with a moderate condition—does the physician apply best practices from studies with patients who have extreme forms of the illness? What about conflicting needs, such as a patient with a non-extreme kidney disease? The best practice is to minimize protein, ideally no more than 4 ounces a day, but that is for more severely ill patients; this patient is also advised to counteract steroid treatment by building up muscle and—you guessed it—ingesting more protein. These kinds of decisions do not line up neatly with EBM and best practices. They require judgment and expertise and processes well studied in cognitive engineering.

The cognitive engineering community can complement the best-practices movement. It seeks to identify subtle aspects of expertise needed to integrate knowledge and apply it to the variability and complexity of individual patient cases, for example, those involving tacit knowledge (e.g., G. Klein, 2009). And it addresses the limitations of rules, procedures, and best practices (Suchman, 1987). The six cognitive challenges we examine in this article illustrate the importance of expertise and the boundary conditions for evidence and best practices.

Therefore, the cognitive engineering community is in a position to develop methods for improving decision making and patient care by taking advantage of evidence without becoming trapped by accepted best practices that cannot cope with complexity (Cook et al., 2000). The cognitive engineering research suggests eight directions for strengthening best practices strategies.

(a) *Develop and sustain expertise.* In order to maintain a balance of evidence and expertise, the cognitive engineering/naturalistic decision making (NDM) community has advocated for ways to build expertise (Ericsson, Charness, Feltovich, & Hoffman, 2006; Klein, 2005; Hoffman et al., 2014). Ericsson (2004) found that the longer it had been since general physicians had completed their formal medical training, the greater the reductions in accuracy and consistency of cardiac diagnoses. That is why it is so important to apply methods for training and sustaining skilled performance in areas such as health care, aviation, intelligence analysis, and so forth. Expertise often takes the form of tacit knowledge, so training in perceptual skills, pattern recognition, anomaly detection, and mental models becomes critical. Decision makers cannot rely on explicit knowledge (facts, rules, procedures) alone. One example of how development of tacit knowledge is being supported is through the ShadowBox training (Klein & Borders, in press), which focuses on speeding the development of expertise through comparisons and introspection.

(b) *Support adaptation.* Expertise depends on tacit knowledge and, in particular, the use of tacit knowledge to judge when to change treatments and when to depart from best practices for individual cases. For example, in health care, we suggest replacing global prescriptions about populations with expectations that physicians will adapt treatment plans to the needs of specific patients. Physicians can shift to customized and annotated plans, not only about medication management but generally about patient treatment strategies.

(c) *Combine evidence with experience.* We see little benefit in contrasting evidence and experience. Both provide a basis for practitioners to make decisions. Sometimes they will conflict, and the decision maker needs to navigate through the conflict. Balancing the competing claims of evidence and experience, and the strengths and limitations of each approach, is not unique to medicine. This duality appears in many different disciplines. For example, the new Pearson Q interactive assessment tool allows clinicians to select tests to be given

(Delis, 2014) but makes suggestions based on other experts (Eric Saperstein, personal communication, January 14, 2015). Experienced behavioral researchers may review the results of a study in terms of means and standard deviations but then look at individual participants for signs of anomalies and unusual patterns.

(d) *Balance generic evidence with experiential evidence.* There is a further conflict within evidence-based approaches: the confidence placed in general evidence drawn from populations versus experiential evidence drawn from the individual cases. In health care, EBM encourages the medical community to rely on generic evidence rather than the evidence of their own patients. However, both types of evidence seem important in making treatment decisions. There are several ways to address this. Gigerenzer (2002) has used frequency data to present generic evidence, with the goal of contextualizing the generic evidence. Another approach is to make better use of displays that illustrate the various parameters and context of a situation to allow for greater resilience (Hollnagel, Woods, & Leveson, 2006; Nemeth, O'Connor, Klock, & Cook, 2006).

(e) *Represent evidence.* Publication of scientific papers is not sufficient. We need to provide data in more easily digested forms to help decision makers see how to personalize the findings to specific cases. Additionally, we advocate ways to offer clearer presentation of effect sizes and clearer presentation of variability, even speculating about the clusters of study participants who gained the most and the least. Using the language of statisticians, we are advocating for ways to highlight Subject \times Treatment interactions.

(f) *Appraise evidence.* Much of the evidence we believe in today will be discarded in 5 to 10 years' time. Decision makers cannot blindly accept the latest studies. They have to gain skills in judging how much confidence to place in evidence. At least, they have to be able to determine that a small difference, although statistically significant, may not warrant too much confidence. The cognitive engineering/NDM community has encountered challenges about appraising evidence with populations, such as intelligence

analysts, who are always alert to the possibility of accidental, erroneous, or even deceptive data points. Panel operators of petrochemical plants need to constantly be alert to the possibility that sensor data may be erroneous. They are trained on garden-path scenarios involving flawed evidence and mistaken initial assessments.

(g) *Share evidence.* Expanding the use of information-exchange mechanisms will help the entire medical community learn about new treatments that have not yet been vetted by the standards of best practices. Several communities have made powerful use of information exchanges, for example, the informal and impromptu lessons learned from chat rooms that sprang up during Operation Iraqi Freedom to trade observations about topics such as detecting improvised explosive devices. The health care domain has established the Patient-Centered Outcomes Research Institute, tasked with sifting data from electronic medical records to identify promising therapies that could make a valuable contribution without having to rely on double-blind experiments. With appropriate oversight, synthesis, and medical legal protection, such forums would provide an opportunity for a broader dialogue on how to balance best practices and expertise.

(h) *Support collaborative decision making.* Cognitive engineering/NDM researchers pay a great deal of attention to effective teamwork. Best practices should not be narrowly drawn as the province of the decision maker. The decision maker has to establish trade-offs and coordination with team members. For example, within health care, the team includes physicians, nurses, and various other professionals. The team concept should be broadened to include patients. A best practice has little value if the patient is unable or unwilling to adhere to the regimen. Physicians can blame the patient, but a more useful stance is to take the patient's abilities and motivations into account in designing a treatment program, even if it means departing from the generally accepted best practices. A suboptimal regimen that a patient can sustain may be better than an optimal one that the patient will ignore. Health care professionals can use what we have learned about adherence (e.g., D. Klein,

2009) in designing individual treatment programs.

CONCLUSIONS

Best practices are an important opportunity for any community to shed outmoded traditions and unreliable anecdotal procedures. They provide an opportunity for scrutiny and debate and progress. They enable organizations to act in a consistent way. However, as we have argued, best practices come with their own challenges.

Cognitive engineering and NDM studies have shown some of the difficulties of using evidence in situations that have a great deal of variability, uncertainty, and risk. In effect, decision makers in domains such as health care need plans like best practices but also need to be effective at revising plans to fit the dynamics and variability of specific situations (e.g., patients and diseases) and to handle the changing knowledge about what is effective.

The approach of cognitive engineering and NDM focuses on layering best practices with experiential knowledge of different situations. In this way, decision makers can handle specific situations, regardless of variability, uncertainty, and change.

We should regard best practices as provisional, not optimal, as a floor rather than a ceiling. When we label an approach a best practice, it tends to become a ceiling that is hard to change even as more knowledge is gained. Instead, we can identify provisional best practices that serve as a floor while learning goes forward. It is a move from "best practices" to "better practices" that frees us from undocumented anecdotal approaches and forces a commitment to continual improvement.

ACKNOWLEDGMENTS

We would like to thank Emilie Roth, Emily Patterson, and Laura Militello for their helpful feedback. We would also like to thank three anonymous reviewers for their extremely thoughtful comments and suggestions.

REFERENCES

- Abbott, K., McKenney, D., & Railsback, P. (2013). *Operational use of flight path management systems*. Retrieved from http://www.faa.gov/about/office_org/headquarters_offices/

- avs/offices/afs/afs400/parc/parc_reco/media/2013/130908_PARC_FltDAWG_Final_Report_Recommendations.pdf
- Amalberti, R. (2013). *Navigating safety: Necessary compromises and tradeoffs, theory and practice*. Dordrecht, Netherlands: Springer-Verlag.
- Brehmer, B. (1987). Development of mental models for decision in technological systems. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error* (pp. 111–120). Chichester, UK: Wiley.
- Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*, 211–241.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.
- Chatterjee, T. (2015). Out of the darkness. *Science*, *350*, 372–375.
- Cook, R. I. (2006). Being bumpable: Consequences of resource saturation and near-saturation for cognitive demands on ICU practitioners. In D. D. Woods & E. Hollnagel (Eds.), *Joint cognitive systems: Patterns in cognitive systems engineering* (pp. 23–35). Boca Raton, FL: CRC Press.
- Cook, R. I., Render, M. L., & Woods, D. D. (2000). Gaps in the continuity of care and progress on patient safety. *British Medical Journal*, *320*, 791–794.
- DeAnda, A., & Gaba, D. (1991). The role of experience in the response to simulated critical incidents. *Anesthesia and Analgesia*, *72*, 308–315.
- Delis, D. (2014, October 31). Cognitive assessment leaps into the digital age. *ESchool News*. Retrieved from <http://www.eschoolnews.com/2014/10/31/cognitive-assessment-digital-429/>
- Denrell, J. (2003). Vicarious learning, undersampling of failure, and the myths of management. *Organization Science*, *14*, 228–243.
- Denrell, J., & Fang, C. (2010). Predicting the next big thing: Success as a signal of poor judgment. *Management Science*, *56*, 1653–1667.
- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *10*, S70–S81.
- Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.). (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge, UK: Cambridge University Press.
- Gaba, D., Maxwell, M., & DeAnda, A. (1987). Anesthetic mishaps: Breaking the chain of accident evolution. *Anesthesiology*, *66*, 670–676.
- Ghaffarzadegan, N., Epstein, A. J., & Martin, E. G. (2013). Practice variation, bias, and experiential learning in Cesarean delivery: A data-based system dynamics approach. *Health Services Research*, *48*, 713–734.
- Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York, NY: Simon & Schuster.
- Gray, J. A. M. (1996). *Evidence-based healthcare*. London, UK: Churchill Livingstone.
- Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York, NY: Psychology Press.
- Hollnagel, E., Woods, D. D., & Leveson, N. (2006). *Resilience engineering: Concepts and precepts*. Farnham, UK: Ashgate.
- Kahneman, D., & Klein, G. A. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*, 515–526.
- Kaiser, J. (2015). The cancer test: A nonprofit's effort to replicate 50 top cancer papers is shaking up labs. *Science*, *348*, 1411–1413.
- Klein, D. E. (2009). The forest and the trees: An integrated approach to designing adherence interventions. *Australasian Medical Journal*, *1*, 181–184.
- Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.
- Klein, G. (2005). *The power of intuition*. New York, NY: Currency/Doubleday.
- Klein, G. (2007a). Flexexecution as a paradigm for replanning, Part 1. *IEEE Intelligent Systems*, *22*, 79–83.
- Klein, G. (2007b). Flexexecution, Part 2: Understanding and supporting flexible execution. *IEEE Intelligent Systems*, *22*, 108–112.
- Klein, G. (2009). *Streetlights and shadows: Searching for the keys to adaptive decision making*. Cambridge, MA: MIT Press.
- Klein, G., & Borders, J. (in press). The ShadowBox approach to cognitive skills training: An empirical evaluation. *Journal of Cognitive Decision Making*.
- Klein, G., Pliske, R., Crandall, B., & Woods, D. (2005). Problem detection. *Cognition, Technology, and Work*, *7*, 14–28.
- Kylesten, B. (2013). Dynamic decision-making on an operative level: A model including preconditions and working method. *Cognitive Technology & Work*, *15*, 197–205.
- Marshall, B. J. (2005, December). *Helicobacter connections*. Nobel lecture, Stockholm, Sweden.
- Nemeth, C., O'Connor, M., Klock, P. A., & Cook, R. (2006). Discovering healthcare cognition: The use of cognitive artifacts to reveal cognitive work. *Organization Studies*, *27*, 1011–1035.
- O'Sullivan, G. C. (2010). Advancing surgical research in a sea of complexity. *Annals of Surgery*, *252*, 711–714.
- Pascale, R., & Sternin, J. (2010). *The power of positive deviance: How unlikely innovators solve the world's toughest problems*. Boston, MA: Harvard Business Review Press.
- Perry, S. J., & Wears, R. L. (2011). Large scale coordination of work: Coping with complex chaos within healthcare. In K. L. Mosier & U. Fisher (Eds.), *Informed by knowledge: Expert performance in complex situations* (pp. 55–59). New York, NY: Taylor & Francis.
- Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., Sexton, B., Hyzy, R., Welsh, R., Roth, G., Bander, J., Kepros, J., & Goeschel, C. (2006). An intervention to decrease catheter-related bloodstream infections in the ICU. *New England Journal of Medicine*, *355*, 2725–2732.
- Rittel, H. W. J., & Weber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*, 155–169.
- Roberts, A. R., & Yeager, K. R. (Eds.). (2004). *Evidence-based practice manual: Research and outcome measures in health and human services*. New York, NY: Oxford University Press.
- Rudolph, J. W., Morrison, J. B., & Carroll, J. S. (2009). The dynamics of action-oriented problem solving: Linking interpretation and choice. *Academy of Management Review*, *34*, 733–756.
- Staszewski, J. (2004). Models of expertise as blueprints for cognitive engineering: Applications to landmine detection. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 458–462). Santa Monica, CA: Human Factors and Ergonomics Society.
- Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, UK: Cambridge University Press.
- Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in healthcare*. Philadelphia, PA: Temple University Press.
- Wears, R. L., & Hunte, G. S. (2014). Seeing patient safety “like a state.” *Safety Science*, *67*, 50–57.

- Wears, R. L., & Schubert, C. C. (2015). Visualizing expertise in context. *Annals of Emergency Medicine*. Advance online publication. doi:10.1016/j.annemergmed.2015.11.027
- Woods, D. D. (1994). Cognitive demands and activities in dynamic fault management. In N. Stanton (Ed.), *Human factors in alarm design* (pp. 63–92). London, UK: Taylor & Francis.
- Woods, D. D. (2005). Creating foresight: Lessons for resilience from Columbia. In W. H. Starbuck & M. Farjoun (Eds.), *Organization at the limit: NASA and the Columbia disaster* (pp. 289–308). Malden, MA: Blackwell.
- Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. Boca Raton, FL: Taylor & Francis.
- Woods, D. D., Roth, E. M., & Bennett, K. B. (1990). Explorations in joint human-machine cognitive systems. In S. Robertson, W. Zachary, & J. Black (Eds.), *Cognition, computing and cooperation* (pp. 123–158). Norwood, NJ: Ablex.
- Zelik, D., Patterson, E. S., & Woods, D. D. (2010). Measuring attributes of rigor in information analysis. In E. S. Patterson & J. Miller (Eds.), *Macro-cognition metrics and scenarios: Design and evaluation for real-world teams* (pp. 65–83). Aldershot, UK: Ashgate.
- Devorah E. Klein, a senior scientist with Marimo Consulting, LLC, is a cognitive psychologist working to design medical products, systems, and services.
- David D. Woods is a professor in the Department of Integrated Systems Engineering at The Ohio State University and past president of the Human Factors and Ergonomics Society and the Resilience Engineering Association.
- Gary Klein is a senior scientist with MacroCognition, LLC, and the author of *Seeing What Others Don't: The Remarkable Ways We Gain Insights*.
- Shawna J. Perry is an emergency medicine physician and visiting scholar at the University of Florida School of Medicine.