Commentary: **Reducing Diagnostic Errors: Another Role for Checklists?**

Bradford D. Winters, PhD, MD, Monica S. Aswani, MSPH, and Peter J. Pronovost, MD, PhD

Abstract

Diagnostic errors are a widespread problem, although the true magnitude is unknown because they cannot currently be measured validly. These errors have received relatively little attention despite alarming estimates of associated harm and death. One promising intervention to reduce preventable harm is the checklist. This intervention has proven successful in aviation, in which situations are linear and deterministic (one alarm goes off and a checklist guides the flight crew to evaluate the cause). In health care, problems are multifactorial and complex. A checklist has been used to reduce central-line-associated bloodstream infections in intensive care units. Nevertheless, this checklist was incorporated in a culture-based safety program that engaged and changed behaviors and used robust measurement of infections to evaluate progress. In this issue, Ely and colleagues describe how three checklists could reduce the cognitive biases and mental shortcuts that underlie diagnostic errors, but point out that these tools still need to be tested. To be effective, they must reduce diagnostic errors (efficacy) and be routinely used in practice (effectiveness). Such tools must intuitively support how the human brain works, and under time pressures, clinicians rarely think in conditional probabilities when making decisions. To move forward, it is necessary to accurately measure diagnostic errors (which could come from mapping out the diagnostic process as the medication process has done and measuring errors at each step) and pilot test interventions such as these checklists to determine whether they work.

Editor's Note: This is a commentary on Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. Acad Med. 2011;86:307–313.

Diagnostic errors are a widespread problem, the magnitude of which is largely unknown because we do not have a valid mechanism to measure such errors. An estimated 80,000 people die annually because of diagnostic errors, 90,000 have a misdiagnosed breast biopsy, and up to 17% of hospitalized patients suffer a diagnostic error (by comparison, about 7% suffer a medication error).[1] One author (B.W.)

studied autopsy reports of intensive care unit (ICU) patients at his institution and found that 43% of patients died from causes not considered or addressed in the care team's treatment plan. Of these, 13% would have been managed differently and experienced different outcomes if they had been accurately diagnosed (Bradford Winters, PhD, MD, unpublished data, November 2010). The scope of the problem is significant, yet diagnostic errors have received relatively little attention from the public and even less attention from research funders.[2] Diagnostic error was mentioned twice in the Institute of Medicine's influential report, *To Err Is Human*, whereas medication error was mentioned 70 times.[2]

Given the alarming estimates described above, why does the problem receive so little attention? It is largely because a misdiagnosis is usually invisible. We cannot measure diagnostic errors fully (autopsies occur in select patients and tell us nothing about nonlethal diagnostic errors), preventing us from making valid and reliable national estimates of those harmed by diagnostic errors and from evaluating the effectiveness of interventions. Unfortunately, there is very little funding for developing the "basic science" of patient safety, and this is especially true for diagnostic errors. If we are to make progress in reducing

harm to patients, we must devote more resources to these critical areas of study.

One promising intervention to reduce preventable harm is the checklist. Checklists were first introduced in aviation in 1935 to mitigate the risks of flying the Boeing B299 bomber; today they are routinely used throughout the aviation industry. Flying an airplane without the use of this important cognitive safety tool is inconceivable. Our research team borrowed from aviation and applied an evidence-based checklist and nearly eliminated central-line-associated bloodstream infections (CLABSIs) in surgical ICUs at the Johns Hopkins Hospital.[3] Use of this simple five-item checklist incorporated in a culture-based safety program was associated with a nearly 70% reduction in the rate of CLABSIs in over 100 ICUs in Michigan, and the results persisted for three years.[4] The state of Rhode Island implemented the same program and achieved similar results.[5] With support from the Agency for Healthcare Research and Quality and private philanthropy, we are implementing this program state by state across the United States.

Although the program used in Michigan is commonly described solely as implementation of a checklist, it is much more complicated than ticking off tasks.

Dr. Winters is assistant professor, Department of Anesthesiology and Critical Care Medicine, and Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland.

Ms. Aswani is Lister Hill Health Policy Fellow, Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.

Dr. Pronovost is professor, Departments of Anesthesiology and Critical Care Medicine, Surgery, and Health Policy and Management, Johns Hopkins University School of Medicine, School of Nursing, and Bloomberg School of Public Health, Baltimore, Maryland.

Correspondence should be addressed to Dr. Pronovost, Johns Hopkins Quality and Safety Research Group, 1909 Thames Street, Second Floor, Baltimore, MD 21231; telephone: (410) 502-3231; fax: (410) 502-3235; e-mail: ppronovo@jhmi.edu.

It requires dramatic behavioral and culture change and robust measurement of infections to realize its full benefit. Culture and behavior change is a difficult and messy process. The checklist itself is only one part of the intervention. The program has several components: rigorous measurement and feedback of infection rates, the Comprehensive Unit-Based Safety Program to improve culture and teamwork,[6] the checklist of five evidence-based practices to prevent CLABSI, and strategies to encourage clinicians (nurses, doctors, and others) to identify and mitigate local barriers to completing the checklist items. Each of these components is needed to substantially reduce infections.[7,8]

In this issue of *Academic Medicine*, Ely and colleagues,[9] pioneers in the field of diagnostic errors and cognitive biases leading to errors, describe how checklists could be applied to the diagnostic process to avoid errors. They argue that checklists can reduce the cognitive biases and mental shortcuts that underlie diagnostic errors. In most diagnostic errors, they perceive that clinicians simply do not think of the diagnosis; therefore, a checklist may help trigger the memory.

The authors present three types of checklists for this purpose. A general checklist asks clinicians whether they obtained a complete and thorough history and physical and whether they reflected on cognitive biases, such as premature closure. A differential diagnosis checklist presents potential diagnoses based on the presence of specific signs and symptoms, such as sinus tachycardia. A "forcing function" checklist presents critical diagnostic approaches to avoid common pitfalls in recognizing specific diseases. The authors present a thoughtful discussion of how these may be used and how they would manage and minimize the required number of checklists.

The authors are thorough in describing how they developed the checklists, but they did not test them in a formal and rigorous manner. Such a test should be undertaken and must look at two key components. Did the checklists reduce diagnostic errors (efficacy studies), and will clinicians use them in routine practice to reduce errors (effectiveness)? Medicine has a clearinghouse of guidelines that are infrequently used.

Tools to improve safety must intuitively support how our brains work rather than how we would like them to work. Under time pressures (the norm for clinicians), clinicians do not make decisions by thinking in conditional probabilities ("if this, then that" statements).[10] Such an approach is laborious and inefficient given time restrictions. Rather, we stick our head in a data stream and recognize patterns or deviations from expected patterns. Efforts to make those patterns or deviations visible may support decision making.[11]

The premise underlying the authors' approach toward building checklists that remind clinicians to do a history and physical or to think of a pulmonary embolus when someone has isolated sinus tachycardia is to reduce errors. Although they may possibly reduce errors, what is the likelihood that reminders and diagnostic guides will actually do so? In reflecting on our own diagnostic errors, we would likely recall that we did do a history, we just did it poorly, or we thought about a pulmonary embolus, but dismissed it. The checklist doesn't present a strong defense against these types of problems, and it may not be a robust and sufficiently comprehensive solution on its own to have the kind of impact we seek. Yet if, as the authors suggest, the most common mistake is simply not thinking of a diagnosis, checklists may help.

Checklists most likely are effective in aviation because they are used during linear and deterministic situations. For instance, when something goes wrong in flight, one alarm goes off at a time, and the checklist guides the flight crew in evaluating the cause of that alarm. In health care, problems are multifactorial and complex. Multiple alarms are frequent and they overlap and interact, especially in environments like the ICU where tachycardia, shortness of breath, hypoxia, and hypertension may all occur simultaneously, each generating an alarm and each requiring a different checklist. Additionally, the predictive value of any particular alarm to indicate a true problem is variable depending on the threshold and patient characteristics. For example, an oxygen saturation of 88% in a patient with advanced chronic obstructive pulmonary disease may be perfectly acceptable, but is clearly not in an otherwise healthy patient. Taking a

history and physical can be viewed as a diagnostic test in which the sensitivity and specificity vary, depending on how skillfully the history and physical are performed. For example, when someone presents with a wide complex tachycardia, asking the patient if he or she has coronary artery disease will have a lower sensitivity than asking whether he or she had a prior myocardial infarction. We can conduct the history and physical, yet miss pulling out key information from the patient. The inexorable increase in reported wrong-site surgeries, despite extensive efforts to prevent them, should cause us to pause. In most of these adverse events, clinicians reportedly used the time-out checklist that was designed to prevent them.

We know a lot about cognitive biases in decision making. We also know that teams make wise decisions because of diverse input and a balance of interdependent discussion with independent voting. Honeybees are a prime example of such group decision making.[11] They also alternate between convergent and divergent thinking,[12] building in pause points to evaluate whether new data fit the mental model.

So how do we move forward in addressing the prevalence of diagnostic errors? First, we need the methodology to accurately measure diagnostic errors (both lethal and nonlethal), providing a national estimate of their impact and evaluating the effectiveness of interventions. Similar to our mapping of medication errors as a process of prescribing, dispensing, administering, and documenting, so too should we view diagnostic errors as a process that include (1) generating a differential diagnosis by taking a history and physical, (2) selecting appropriate tests to narrow that differential, (3) conducting the tests appropriately to obtain reliable data, (4) interpreting the results correctly with minimal bias, (5) selecting the working diagnosis, and (6) deciding on the appropriate treatment and follow-up to reexamine whether the working diagnosis is in fact still supported by the clinical evidence (response to treatment, change in symptoms) in an ongoing fashion. Errors occurring at each of these steps could be measured.

Second, we should pilot test interventions, such as the proposed

checklists, in a rigorous way to determine both their efficacy and effectiveness. Drs. Ely, Graber, and Croskerry are leading the way. Researchers need to be mindful of how the guidelines are used and not used in practice, including the barriers and benefits of the guidelines. After all, health care is replete with elegant, scholarly, but infrequently used guidelines.

Third, we need to invest in the "basic science" of diagnostic errors to more deeply understand why we make mistakes and how we can prevent them. This will require interdisciplinary teams. Clinicians, cognitive psychologists, human-factors and system engineers, behavioral economists, organizational sociologists, health services researchers, and informatics experts will all have a role in this task.

Diagnostic errors result in significant preventable harm, yet for too long, these errors have gone unnoticed. The article by Ely and colleagues seeks to change that. They are trying a novel yet practical intervention— checklists. Let's hope physicians are humble enough to give them a fair trial in the effort to reduce diagnostic errors and the harm they cause.

References

1 Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: A systematic review. JAMA. 2003;289:2849–2856.

2 Wachter RM. Why diagnostic errors don't get any respect—and what can be done about them. Health Aff (Millwood). 2010;29:1605–1610.

3 Berenholtz SM, Pronovost PJ, Lipsett PA, et al. Eliminating catheter-related bloodstream infections in the intensive care unit. Crit Care Med. 2004;32:2014–2020.

4 Pronovost PJ, Goeschel CA, Colantuoni E, et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: Observational study. BMJ. 2010; 340:c309.

5 DePalo VA, McNicoll L, Cornell M, Rocha JM, Adams L, Pronovost PJ. The Rhode Island ICU collaborative: A model for reducing central line-associated bloodstream infection and ventilator-associated pneumonia statewide. Qual Saf Health Care. 2010; 19:555–571.

6 Timmel J, Kent PS, Holzmueller CG, Paine LA, Schulick RD, Pronovost PJ. Impact of the comprehensive unit-based safety program (CUSP) on safety culture in a surgical inpatient unit. Jt Comm J Qual Patient Saf. 2010;36:252–260.

7 Bosk CL, Dixon-Woods M, Goeschel CA, Pronovost PJ. Reality check for checklists. Lancet. 2009;374:444–445.

8 Pronovost PJ. Learning accountability for patient outcomes. JAMA. 2010;302:204–205.

9 Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. Acad Med. 2011; 86:307–313.

10 Klein G. Sources of Power: How People Make Decisions. Cambridge, Mass: Massachusetts Institute of Technology; 1999.

11 Klein GA. The Power of Intuition: How to Use Your Gut Feelings to Make Better Decisions at Work. 2nd ed. New York, NY: Crown Business; 2004.

12 Tetlock PE. Expert Political Judgement. Princeton, NJ: Princeton University Press; 2005.