



Minimizing Diagnostic Error: The Importance of Follow-up and Feedback

An open-loop system (also called a “nonfeedback controlled” system) is one that makes decisions based solely on preprogrammed criteria and the preexisting model of the system. This approach does not use feedback to calibrate its output or determine if the desired goal is achieved. Because open-loop systems do not observe the output of the processes they are controlling, they cannot engage in learning. They are unable to correct any errors they make or compensate for any disturbances to the process. A commonly cited example of the open-loop system is a lawn sprinkler that goes on automatically at a certain hour each day, regardless of whether it is raining or the grass is already flooded.¹

To an unacceptably large extent, clinical diagnosis is an open-loop system. Typically, clinicians learn about their diagnostic successes or failures in various ad hoc ways (e.g., a knock on the door from a server with a malpractice subpoena; a medical resident learning, upon bumping into a surgical resident in the hospital hallway that a patient he/she cared for has been readmitted; a radiologist accidentally stumbling upon an earlier chest x-ray of a patient with lung cancer and noticing a nodule that had been overlooked). Physicians lack systematic methods for calibrating diagnostic decisions based on feedback from their outcomes. Worse yet, organizations have no way to learn about the thousands of collective diagnostic decisions that are made each day—information that could allow them to both improve overall performance as well as better hear the voices of the patients living with the outcomes.²

THE NEED FOR SYSTEMATIC FEEDBACK

In this commentary, I consider the issues raised in the review by Drs. Berner and Graber³ and take the discussion further in contemplating the need for systematic feedback to improve diagnosis. Whereas their emphasis centers around

the question of physician overconfidence regarding their own cognitive abilities and diagnostic decisions, I suspect many physicians feel more beleaguered and distracted than overconfident and complacent. There simply is not enough time in their rushed outpatient encounters, and too much “noise” in the nonspecified undifferentiated complaints that patients bring to them, for physicians, particularly primary care physicians, to feel overly secure. Both physicians and patients know this. Thus, we hear frequent complaints from both parties about brief appointments lacking sufficient time for full and proper evaluation. We also hear physicians’ confessions about excessive numbers of tests being done, “overordered” as a way to compensate for these constraints that often are conflated with and complicated by “defensive medicine”—usually tests and consults ordered solely to block malpractice attorneys.

The issue is not so much that physicians lack an awareness of the thin ice on which they often are skating, but that they have no consistent and reliable systems for obtaining feedback on diagnosis. The reasons for this deficiency are multifactorial. **Table 1** lists some of the factors that mitigate against more systematic feedback on diagnosis outcomes and error. These items invite us to explicitly recognize this problem and design approaches that will make diagnosis more of a closed rather than open-loop system.

Given the current emphasis on heuristics, cognition, and unconscious biases that has been stimulated by publications such as Kassier and Kopelman’s classic book *Learning Clinical Reasoning*,⁴ and *How Doctors Think*,⁵ the recent bestseller by Dr. Jerome Groopman, it is important to keep in mind that good medicine is less about brilliant diagnoses being made or missed and more about mundane mechanisms to ensure adequate follow-up.⁶ Although this assertion remains an untested empirical question, I suspect that the proportion of malpractice cases related to diagnosis error—the leading cause of malpractice suits, outnumbering claims from medication errors by a factor of 2:1—that concern failure to *consider* a particular diagnosis is less than imagined.^{7,8} Despite popular imagery of a diagnosis being missed by a dozen previous physicians only to be eventually made correctly by a virtuoso thinker (such as that stimulated by the Groopman book and dramatic cases reported in the

Statement of Author Disclosure: Please see the Author Disclosures section at the end of this article.

Requests for reprints should be addressed to: Gordon D. Schiff, MD, Division of General Medicine, Brigham and Women’s Hospital, 1620 Tremont, 3rd Floor, Boston, Massachusetts 02120.

E-mail address: gshiff@partners.org.

Table 1 Barriers to feedback and follow-up

- Physician lack of time and systematic approaches for obtaining follow-up
 - Unrealistic to expect MDs to rely on memory or ad hoc methods
- Clinical practice often doesn't require a diagnosis to treat
 - Blunts MDs interest in feedback/follow-up
 - Legitimately seen as purely academic question
 - Suggests it is not worth time for follow-up
- High frequency of symptoms for which no definite diagnosis is ever established
 - Self-limited nature of many symptoms/diagnoses
 - Nonspecific symptoms for which no "organic" etiology ever identified
- Threatening nature of critical feedback makes MDs defensive
 - MDs pride themselves on being "good diagnosticians"
 - Reluctance of colleagues to "criticize" peers and be critiqued by them
- Fragmentation and discontinuities of care
 - Ultimate diagnoses are often made later, in different setting
 - Patient seen in other ERs, by specialists, admitted to different hospital
 - No organized system for feedback of findings across institutions
- Reliance on patient return for follow-up; fragile link
 - Patients busy; inconvenient to return
 - Cost barriers
 - Out-of-pocket costs from first visit can inhibit return
 - Perceived lack of "value" for return visit
 - If improved, seems pointless
 - If not improved, may also seem not worthwhile
 - Patient satisfaction and convenience
 - If not improved, disgruntled patient may seek care elsewhere
- Managed care barriers discourage access
 - Prior approval often required for repeat visit
- "Information breakage" despite return to original setting/MD
 - Original record or question(s) may be inaccessible or forgotten
 - May see partner of MD or other member of team

ER = emergency room; MD = medical doctor.

press), I believe such cases are less common than those involving failure to definitively establish a diagnosis that was considered by one or more physicians earlier. Obvious examples include the case of a patient with chest pain being sent home from the emergency room (ER) with a missed myocardial infarction (MI) or that involving oversight of a subtle abnormality on mammogram. Every ER physician in the emergency considers MI in chest-pain patients, and why else is a mammogram performed other than for consideration of breast cancer?

EXPANDED PARADIGMS IN DIAGNOSIS

The true concern in routine clinical diagnosis is not whether unsuspected new diagnoses are made or missed as much as

it is the complexities of weighing and pursuing diagnostic considerations that are either obvious, may have been previously considered, or simply represent "dropped balls" (e.g., failed follow-up on an abnormal test result).⁹ Furthermore, other paradigms often turn out to be more important than simply affixing a label on a patient naming a specific diagnosis (**Table 2**). Central to each of these "expanded paradigms" is the role for follow-up: deciding when a patient is acutely ill and required hospitalization, versus relatively stable but in need of careful observation, watching for complications or response after a diagnosis is made and a treatment started, monitoring for future recurrences, or even simply revising the diagnosis as the syndrome evolves. It often is more important for an ER or primary care physician to accurately decide whether a patient is "sick" and needs to be hospitalized or sent home than it is to come up with the precisely correct diagnosis at that moment of first encounter.

RESPONSE OVER TIME: THE ULTIMATE TEST?

Although the traditional "test of time" is frequently invoked, it is rarely applied in a standardized or evidence-based fashion, and never in a way that involves systematic tracking and calculating of accuracy rates or formal use of data that evolves over time for recalibration. One key unanswered question is, To what extent can we judge the accuracy of diagnoses based on how patients do over time or respond to treatment? In other words, if a patient gets better and responds to recommended therapy, can we assume the treatment, and hence the diagnosis, was correct? Basing diagnosis accuracy and learning on capturing feedback on whether or not a patient successfully "responds" to treatment is fraught with nuances and complexities that are rarely explicitly considered or measured. A partial list of such complexities is shown in **Table 3**.

Despite these limitations, feedback on patient response is critical for knowing not just how the patient is doing but how we as clinicians are doing. Particularly if we are mindful of these pitfalls, and especially if we can build in rigor with quantitative data to better answer the above questions, feedback on response seems imperative to learning from and improving diagnosis.

VIEWING DIAGNOSIS AS A RELATIONSHIP RATHER THAN A LABEL

Feedback on how patients are doing embodies an important corollary to the entire paradigm of diagnosis tracking and feedback. To a certain extent, diagnosis has been "reified," i.e., taken as an abstraction—an artificially constructed label—and misconceived as a "fact of nature."^{10,11} By turning complex dynamic relationships between patients and their social environments, and even relationships between physicians and their patients, into "things" that boil down to neat categories, we risk oversimplifying complicated interactions of factors that are, in practice, larger than an International Classification of Diseases, 9th Revision (ICD-9) or

Table 2 Limitations of using successful or failed "treatment response" as an indicator for diagnostic error

- Diagnosis of severity/acuity
 - Failure to recognize patient need to be hospitalized or sent to ICU
- Diagnosis of complication
 - Assessing sequelae of a disease, drug, or surgery
- Diagnosis of a recurrence
 - What follow-up surveillance is required and how to interpret results
- Diagnosis of cure or failure to respond
 - When can clinician feel secure vs worry if symptoms don't improve
 - When should "test-of-cure" be done routinely
- Diagnosis of a misdiagnosis
 - When should a previous diagnosis be questioned and revised

ICU = intensive care unit.

Table 3 Factors complicating assessment of treatment response

- Patients who respond to a nonspecific/nonselective drug (e.g., corticosteroids) despite a wrong diagnosis
- Patients who fail to respond to therapy despite the correct diagnosis
- Varying time intervals for expected response
 - When does a clinician decide a patient is/is not responding
- Interpretation of partial responses
- How to incorporate known variations in response
 - Timing
 - Degree
- Role of surrogate (e.g., lab test or x-ray improvement) vs actual clinical outcome
- Timing of repeat testing to check for patient response
 - When and how often to repeat an x-ray or blood test
- Role of mitigating factors
 - Self-limited illnesses
 - Placebo response
 - Naturally relapsing and remitting courses of disorders

Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) label.¹²

Building dialogue into the clinical diagnostic process, whereby the patient tells the practitioner how he/she is doing, represents an important premise. At the most basic level, doing so demonstrates a degree of caring that extends the clinical encounter beyond the rushed 15-minute exam. It is impossible to exaggerate the amazement and appreciation of my patients when I call to ask how they are doing a day or a week after an appointment to follow up on a clinical problem (as opposed to them calling me to complain that they are not improving!). Such follow-up means acknowledging that patients are coproducers in diagnosis—that they have an extremely important role to play to ensure that our diagnoses are as accurate as possible.¹³

The concept of coproduction of diagnosis goes beyond patients going home and "googling" the diagnosis the physi-

cian has suggested in order to decide whether their symptoms are consistent with what they read on the Internet, although there is certainly a role for such searches. It also is about much more than patients obtaining a second opinion from a second physician to enhance and ensure the accuracy of the diagnosis they were given (although this also is happening all the time, and we lack good ways to learn from such error-checking activities). What coproduction of diagnosis really should mean is that the patient is a partner in thinking through and testing the diagnostic hypothesis and has various important roles to play, some of which are described below.

Confirming or refuting a diagnostic hypothesis based on temporal relationships. "Doc, I know you think this rash is from that drug, but I checked and the rash started a week before I began the medication," or "The fever started before I even went to Guatemala."

Noting relieving or exacerbating factors that otherwise might not have been considered. "I later noticed that every time I leaned forward it made my chest pain better." This is a possible clue for pericarditis.

Carefully assessing the response to treatment. "The medication seemed to help at first, but is no longer helping." This suggests that the diagnosis or treatment may be incorrect (see Table 3).

Feeding back the nuances of the comments of a specialist referral. "The cardiologist you sent me to didn't think the chest pain was related to the mitral valve problem but she wasn't sure."

Triggering other past historical clues. "After I went home and thought about it, I remembered that as a teenager I once had an injury to my left side and peed blood for a week," states a patient with an otherwise inexplicable nonfunctioning left kidney. "I remembered that I once did work in a factory that made batteries," offers a patient with an elevated lead level.

Should I, as the physician of each of the actual patients cited above, have "taken a better history" and uncovered each of these pieces of data myself on the initial visit? Each emerged only through subsequent follow-up. Shouldn't I have asked more detailed probing questions during my first encounter with the patient? Shouldn't I have asked follow-up questions during the initial encounter that more actively explored my differential diagnosis based on (what ideally should be) my extensive knowledge of various diseases? Realistically, this will never happen.

Hit-and-miss medicine needs to be replaced by pull systems, which are described by Najarian¹⁴ as "going forward by moving backward." Communication fed back from downstream outcomes, like Japanese kanban cards, should reliably pull the physician back to the patient to adjust

his/her management as well as continuously redesign methods for approaching future patients.

AVOIDANCE OF TAMPERING

Carefully refined signals from downstream feedback represent an important antidote to a well-known cognitive bias, *anchoring*, i.e., fixing on a particular diagnosis despite cues and clues that such persistence is unwarranted. However, feedback can exacerbate another bias—*availability bias*,¹⁵ i.e., overreacting to a recent or vividly recalled event. For example, upon learning that a patient with a headache that was initially dismissed as benign was found to have a brain tumor, the physician works up all subsequent headache patients with imaging studies, even those with trivial histories. Thus, potentially useful feedback on the patient with a missed brain tumor is given undue weight, thereby biasing future decisions and failing to properly account for the rarity of neoplasms as a cause of a mild or acute headache.

When the quality guru Dr. W. Edwards Deming came into a factory, one of the first ways he improved quality was to stop the well-intentioned workers from “tampering,” i.e., fiddling with the “dials.”¹⁶ For example, at the Wausau Paper company, the variations in paper size decreased by simply halting repeated adjustments of the sizing dials, which Deming showed often represented chasing random variation. As he dramatically showed with his classic funnel experiment, in which subjects dropped marbles through a funnel over a bull’s-eye target, the more the subject attempted to adjust the position to compensate for each drop (e.g., moving to the right when a marble fell to the left of the target), the more variation was introduced, resulting in fewer marbles hitting the target than if the funnel were held in a consistent position. By overreacting to this random variation each time the target was missed, the subjects worsened rather than improved their accuracy and thereby were even less likely to hit the target.

If each time a physician’s discovery that his/her diagnostic assessment erred on the side of a making a common diagnosis (thus missing a rare disorder) led to overreactions regarding future patients, or conversely, if each time the physician learned of a fruitless negative workup for a rare diagnosis, he/she vowed never to order so many tests, our cherished continuous feedback loops merely could be adding to variations and exacerbating poor quality in diagnosis. Or to paraphrase the language of Berner and Graber³ or Rudolph,¹⁷ feedback that inappropriately leads to either shaking or bolstering the physician’s confidence in future diagnostic decision making is perhaps doing more harm than good. The continuous quality improvement (CQI) notion of avoiding tampering can be seen as the counterpart to the cognitive availability bias. It suggests a critical need to develop methods to properly weigh feedback in order to better calibrate diagnostic decision making. Although some of the so-called “statistical process control” (SPC) rules can be adapted to ensure more quantitative rigor to recalibrating decisions, generally, physicians are unfamiliar with these

techniques. Thus, developing easy ways to incorporate, weigh, and simplify feedback data needs to be a priority.

CONCLUSION

Learning and feedback are inseparable. The old tools—ad hoc fortuitous feedback, individual idiosyncratic systems to track patients, reliance on human memory, and patient adherence to or initiating of follow-up appointments—are too unreliable to be depended upon to ensure high quality in modern diagnosis. Individual efforts to become wiser from cumulative clinical experience, an uphill battle at best, lack the power to provide the intelligence needed to inform learning organizations. What is needed instead is a systematic approach, one that fully involves patients and possesses an infrastructure this is hard wired to capture and learn from patient outcomes. Nothing less than such a linking of disease natural history to learning organizations poised to hear and learn from patient experiences and physician practices will suffice.

Gordon D. Schiff, MD
Division of General Medicine
Brigham and Women’s Hospital
Boston, Massachusetts, USA

AUTHOR DISCLOSURES

The author reports the following conflicts of interest with the sponsor of this supplement article or products discussed in this article:

Gordon D. Schiff, MD, has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

References

1. Open-loop controller. Available at: http://www.en.wikipedia.org/wiki/Open-loop_controller. Accessed January 23, 2008.
2. Schiff GD, Kim S, Abrams R, et al. Diagnosing diagnostic errors: lessons from a multi-institutional collaborative project. In: *Advances in Patient Safety: From Research to Implementation*, vol 2. Rockville, MD: Agency for Healthcare Research & Quality [AHRQ], February 2005. AHRQ Publication No. 050021. Available at: <http://www.ahrq.gov/qual/advances/>. Accessed December 3, 2007.
3. Berner E, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121(suppl 5A):S2–S23.
4. Kassirer JP, Kopelman RI. *Learning Clinical Reasoning*. Baltimore, MD: Lippincott Williams & Wilkins, 1991.
5. Groopman J. *How Doctors Think*. New York: Houghton Mifflin, 2007.
6. Schiff GD. Commentary: diagnosis tracking and health reform. *Am J Med Qual*. 1994;9:149–152.
7. Phillips R, Bartholomew L, Dovey S, Fryer GE Jr, Miyoshi TJ, Green LA. Learning from malpractice claims about negligent, adverse events in primary care in the United States. *Qual Saf Health Care*. 2004;13:121–126.
8. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med*. 2006;145:488–496.
9. Gandhi TK. Fumbled handoffs: one dropped ball after another. *Ann Intern Med*. 2005;142:352–358.
10. Gould SJ. *The Mismeasure of Man*. New York: Norton & Co, 1981.
11. Freeman A. Diagnosis as explanation. *Early Child Dev Care*. 1989;44:61–72.

12. Mellsop G, Kumar S. Classification and diagnosis in psychiatry: the emperor's clothes provide illusory court comfort. *Psychiatry Psychol Law*. 2007;14:95–99.
13. Hart JT. *The Political Economy of Health Care. A Clinical Perspective*. Bristol, United Kingdom: The Policy Press, 2006.
14. Najarian G. The pull system mystery explained: drum, buffer and rope with a computer. The Manager.org. Available at: http://www.themanager.org/strategy/pull_system.htm. Accessed January 24, 2008.
15. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science*. 1974;185:1124–1130.
16. Deming WE. *Out of the Crisis*. Cambridge, MA: MIT Press, 1982.
17. Rudolph JW. Confidence, error, and ingenuity in diagnostic problem solving: clarifying the role of exploration and exploitation. Presented at: Annual Meeting of the Healthcare Management Division of the Academy of Management. August 5–8, 2007; Philadelphia, PA.