# THE AMERICAN JOURNAL *of* MEDICINE®

May 2008
Volume 121 (5A)


# Diagnostic Error: Is Overconfidence the Problem?


**GUEST EDITORS**

**Mark L. Graber, MD, FACP**
Chief, Medical Service
Veterans Affairs Medical Center
Northport, New York
Professor and Associate Chair
Department of Medicine
SUNY Stony Brook
Stony Brook, New York

**Eta S. Berner, EdD, FACMI, FHIMSS**
Professor, Health Informatics
Department of Health Services Administration
School of Health Professions
University of Alabama at Birmingham
Birmingham, Alabama



This supplement was sponsored by the Paul Mongerson Foundation through the Raymond James Charitable Endowment Fund. Many of the ideas expressed here emerged from discussions at a meeting among the authors in Naples, Florida, in December 2006 that was sponsored by the University of Alabama at Birmingham with support from the Paul Mongerson Foundation.


**Statement of Peer Review:** All supplement manuscripts submitted to *The American Journal of Medicine* for publication are reviewed by the Guest Editor(s) of the supplement, by an outside peer reviewer who is independent of the supplement project, and by the Journal's Supplement Editor (who ensures that questions raised in peer review have been addressed appropriately and that the supplement has an educational focus that is of interest to our readership).

**Author Disclosure Policy:** All authors contributing to supplements in *The American Journal of Medicine* are required to fully disclose any primary financial relationship with a company that has a direct fiscal or financial interest in the subject matter or products discussed in the submitted manuscripts, or with a company that produces a competing product. These relationships (e.g., ownership of stock or significant honoraria or consulting fees) and any direct support of research by a commercial company must be indicated on the title page of each manuscript. This information will be published in the frontmatter of each supplement.

The American Journal of Medicine (2008) Vol 121 (5A), fvii

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

# Foreword

After being misdiagnosed with pancreatic cancer in 1980, I founded the Computer Assisted Medical Diagnosis and Treatment Foundation to improve the accuracy of medical diagnosis. The foundation has sponsored programs to develop and evaluate computerized programs for medical diagnosis and to encourage physicians to use computers for their order entry. My role was insignificant, but as the result of much work by many people, substantial progress has been made. Physicians today are clearly more accepting of computer assistance and this movement is accelerating.

However, in 2006, I became worried after questioning my personal physicians as to why they did not use computers for diagnosis more often. Most explained that their diagnostic error rate was $<1\%$ and that computer use was time consuming. However, I had read that studies of diagnostic problem solving showed an error rate ranging from 5% to 10%. The physicians attributed the higher error rates to "other" less skilled physicians; few felt a need to improve their own diagnostic abilities.

From my perspective as a patient, even an error rate of 1% is unacceptable. It is ironic that most physicians I have asked are convinced there is much room for improvement in diagnosis—by other physicians. In my view, diagnostic error will be reduced only if physicians have a more realistic understanding of the amount of diagnostic errors they personally make. I believe that the accuracy of diagnosis can be best improved by informing physicians of the extent of their own (not others') errors and urging them to personally take steps to reduce their own mistakes.

It is logical that physicians' overconfidence in their ability inadvertently reduces the attention they give to reducing their own diagnostic errors. Unfortunately, this sensitive problem is rarely discussed and it is understudied. This supplement to *The American Journal of Medicine*, which features Drs. Eta S. Berner and Mark L. Graber's comprehensive review of a broad range of literature on the extent of diagnostic errors, the causes, and strategies to reduce them, addresses that gap.

Drs. Berner and Graber conducted the literature review and developed a framework for strategies to address the problem. Their colleagues' commentaries expand and refine our understanding of the causes of errors and the strategies to reduce them. The papers in this supplement confirm the extent of diagnostic errors and suggest improvement will best come by developing systems to provide physicians with better feedback on their own errors.

Hopefully this set of articles will inspire us to improve our own diagnostic accuracy and to develop systems that will provide diagnostic feedback to all physicians.

Paul Mongerson, BSME
*From the Paul Mongerson Foundation within the Raymond James Charitable Endowment Fund*

Requests for reprints should be addressed to: 7425 Pelican Bay Boulevard, Apartment 703, Naples, FL 34108.

E-mail address: pmongerson@aol.com


doi:10.1016/j.amjmed.2008.02.008

The American Journal of Medicine (2008) Vol 121 (5A), S1

THE AMERICAN JOURNAL *of* MEDICINE®

# Introduction

This supplement to *The American Journal of Medicine* centers on the widely acknowledged occurrence of frequent errors in medical practice, especially in medical diagnosis. In the featured article, Drs. Eta S. Berner and Mark L. Graber bring our attention directly to the paucity of penitents among the crowd of seemingly unaware sinners. They convincingly demonstrate that we physicians lack strong direct and timely feedback about our decisions. Given that most medical decisions, however curious our reasoning, actually work relatively well within our chosen practice situation, we are not acutely anxious about oversights. In other words, the average day does not confront us with our errors.

Drs. Berner and Graber summarize an extensive body of scholarly writing about teaching, learning, reasoning, and decision making as it relates to diagnostic error and overconfidence, which is expanded upon by their colleagues. In the first commentary, Drs. Pat Croskerry and Geoff Norman review 2 modes of clinical reasoning in an effort to better understand the processes underlying overconfidence. Ms. Beth Crandall and Dr. Robert L. Wears highlight gaps in knowledge about the nature of diagnostic problems, emphasizing the limitations of applying static models to the messy world of clinical practice. Clearly, many experts are concerned about these processes. I commend this volume to any professional or lay reader who thinks it is easy to bring medical decision making closer to the ideal.

One finds a theme repeating in these carefully reasoned papers: namely, that, as phrased by Dr. Gordon L. Schiff in the fourth commentary, "Learning and feedback are inseparable." This issue is addressed from a variety of perspectives. In the third commentary, Drs. Jenny W. Rudolph and J. Bradley Morrison provide an expanded model of the fundamental feedback processes involved in diagnostic problem solving, highlighting particular leverage points for avoiding error. Dr. Schiff explicates the numerous barriers to adequate feedback and follow-up in the real world of clinical practice and emphasizes the need for a systematic tracking approach over time that fully involves patients. In the final commentary, Dr. Graber identifies stakeholders interested in medical diagnosis and provides recommendations to help each reduce diagnostic error.

These papers sound a second theme, also worth noting. That is, medical practitioners really do not use systems designed to aid their diagnostic decision making. The exception is the case already recognized to be miserably complex or misdiagnosed! This fits my own experience. In the 1980s, I developed a system to aid medical reasoning called CONSIDER. Its purpose was to increase the likelihood that the correct diagnosis appeared on the list of differential diagnoses considered by the physician. Although surprisingly apt (and offered free of charge by Missouri Regional Medical Program), the system produced many astonishing and, at times, amusing anecdotal reports, particularly regarding "tough" cases, but no rush to employment or major changes in mortality rates.

Consequently, I sympathize with and respectfully salute these present efforts to study diagnostic decision making and to remedy its weaknesses. In closing, I applaud especially the suggestions to systematize the incorporation of the "downstream" experiences and participation of the patients in all efforts to improve the diagnostic process. These problems likely will not get better until the average day *does* confront us with our errors.

Donald A.B. Lindberg, MD
*Director, National Library of Medicine*
*National Institutes of Health*
*Department of Health and Human Services*
*Bethesda, Maryland, USA*

## AUTHOR DISCLOSURES

**Donald A.B. Lindberg, MD,** has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

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

Requests for reprints should be addressed to Donald A.B. Lindberg, MD, National Library of Medicine, National Institutes of Health, Building 38/Room 2 E17, 8600 Rockville Pike, Bethesda, Maryland 20894.

E-mail address: lindberg@nlm.nih.gov.


doi:10.1016/j.amjmed.2008.02.007

The American Journal of Medicine (2008) Vol 121 (5A), S2–S23

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

# Overconfidence as a Cause of Diagnostic Error in Medicine

**Eta S. Berner, EdD,[a] and Mark L. Graber, MD[b]**

[a]*Department of Health Services Administration, School of Health Professions, University of Alabama at Birmingham, Birmingham, Alabama, USA; and* [b]*VA Medical Center, Northport, New York and Department of Medicine, State University of New York at Stony Brook, Stony Brook, New York, USA*

**ABSTRACT**

The great majority of medical diagnoses are made using automatic, efficient cognitive processes, and these diagnoses are correct most of the time. This analytic review concerns the exceptions: the times when these cognitive processes fail and the final diagnosis is missed or wrong. We argue that physicians in general underappreciate the likelihood that their diagnoses are wrong and that this tendency to overconfidence is related to both intrinsic and systemically reinforced factors. We present a comprehensive review of the available literature and current thinking related to these issues. The review covers the incidence and impact of diagnostic error, data on physician overconfidence as a contributing cause of errors, strategies to improve the accuracy of diagnostic decision making, and recommendations for future research. 

**KEYWORDS:** Cognition; Decision making; Diagnosis; Diagnosis, computer-assisted; Diagnostic errors; Feedback

*Not only are they wrong but physicians are "walking . . . in a fog of misplaced optimism" with regard to their confidence.*

—Fran Lowry[1]

Mongerson[2] describes in poignant detail the impact of a diagnostic error on the individual patient. Large-scale surveys of patients have shown that patients and their physicians perceive that medical errors in general, and diagnostic errors in particular, are common and of concern. For instance, Blendon and colleagues[3] surveyed patients and physicians on the extent to which they or a member of their family had experienced medical errors, defined as mistakes that "result in serious harm, such as death, disability, or additional or prolonged treatment." They found that 35% of physicians and 42% of patients reported such errors.

A more recent survey of 2,201 adults in the United States commissioned by a company that markets a diagnostic decision-support tool found similar results.[4] In that survey, 35% experienced a medical mistake in the past 5 years involving themselves, their family, or friends; half of the mistakes were described as diagnostic errors. Of these, 35% resulted in permanent harm or death. Interestingly, 55% of respondents listed misdiagnosis as the greatest concern when seeing a physician in the outpatient setting, while 23% listed it as the error of most concern in the hospital setting. Concerns about medical errors also were reported by 38% of patients who had recently visited an emergency department; of these, the most common worry was misdiagnosis (22%).[5]

These surveys show that patients report frequent experience with diagnostic errors and/or that these errors are of significant concern for them in their encounters with the healthcare system. However, as pointed out in an editorial by Tierney,[6] patients may not always interpret adverse events accurately, or may differ with their physicians as to the reason for the adverse event. For this reason, we have reviewed the scientific literature on the incidence and impact of diagnostic error and have examined the literature on overconfidence as a contributing cause of diagnostic errors. In the latter portion of this article we review the literature on the effectiveness of potential strategies to reduce diagnostic error and recommend future directions for research.

This research was supported through the Paul Mongerson Foundation within the Raymond James Charitable Endowment Fund (ESB) and the National Patient Safety Foundation (MLG).

*Statement of author disclosures:* Please see the Author Disclosures section at the end of this article.

Requests for reprints should be addressed to Eta S. Berner, EdD, Department of Health Services Administration, School of Health Professions, University of Alabama at Birmingham, 1675 University Boulevard, Room 544, Birmingham, Alabama 35294-3361.

E-mail address: eberner@uab.edu.

0002-9343/$ -see front matter 
doi:10.1016/j.amjmed.2008.01.001

## INCIDENCE AND IMPACT OF DIAGNOSTIC ERROR

We reviewed the scientific literature with several questions in mind: (1) What is the extent of incorrect diagnosis? (2) What percentage of documented adverse events can be attributed to diagnostic errors and, conversely, how often do diagnostic errors lead to adverse events? (3) Has the rate of diagnostic errors decreased over time?

### What is the Extent of Incorrect Diagnosis?

Diagnostic errors are encountered in every specialty, and are generally lowest for the 2 perceptual specialties, radiology and pathology, which rely heavily on visual interpretation. An extensive knowledge base and expertise in visual pattern recognition serve as the cornerstones of diagnosis for radiologists and pathologists.[7] The error rates in clinical radiology and anatomic pathology probably range from 2% to 5%,[8–10] although much higher rates have been reported in certain circumstances.[9,11] The typically low error rates in these specialties should not be expected in those practices and institutions that allow x-rays to be read by frontline clinicians who are not trained radiologists. For example, in a study of x-rays interpreted by emergency department physicians because a staff radiologist was unavailable, up to 16% of plain films and 35% of cranial computed tomography (CT) studies were misread.[12]

Error rates in the clinical specialties are higher than in perceptual specialties, consistent with the added demands of data gathering and synthesis. A study of admissions to British hospitals reported that 6% of the admitting diagnoses were incorrect.[13] The emergency department requires complex decision making in settings of above-average uncertainty and stress. The rate of diagnostic error in this arena ranges from 0.6% to 12%.[14,15]

Based on his lifelong experience studying diagnostic decision making, Elstein[16] estimated that the rate of diagnostic error in clinical medicine was approximately 15%. In this section, we review data from a wide variety of sources that suggest this estimate is reasonably correct.

**Second Opinions and Reviews.** Several studies have examined changes in diagnosis after a second opinion. Kedar and associates,[17] using telemedicine consultations with specialists in a variety of fields, found a 5% change in diagnosis. There is a wealth of information in the perceptual specialties using second opinions to judge the rate of diagnostic error. These studies report a variable rate of discordance, some of which represents true error, and some is disagreement in interpretation or nonstandard defining criteria. It is important to emphasize that only a fraction of the discordance in these studies was found to cause harm.

***Dermatology.*** Most studies focused on the diagnosis of pigmented lesions (e.g., ruling out melanoma). For example, in a study of 5,136 biopsies, a major change in diagnosis was encountered in 11% on second review. Roughly 1% of diagnoses were changed from benign to malignant, roughly 1% were downgraded from malignant to benign, and in roughly 8% the tumor grade was changed enough to alter treatment.[18]

***Anatomic Pathology.*** There have been several attempts to determine the true extent of diagnostic error in anatomic pathology, although the standards used to define an error in this field are still evolving.[19] In 2000, The American Society of Clinical Pathologists convened a consensus conference to review second opinions in anatomic pathology.[20] In 1 such study, the pathology department at the Johns Hopkins Hospital required a second opinion on each of the 6,171 specimens obtained over an 18-month period; discordance resulting in a major change of treatment or prognosis was found in just 1.4 % of these cases.[10] A similar study at Hershey Medical Center in Pennsylvania identified a 5.8% incidence of clinically significant changes.[20] Disease-specific incidences ranged from 1.3% in prostate samples to 5% in tissues from the female reproductive tract and 10% in cancer patients. Certain tissues are notoriously difficult; for example, discordance rates range from 20% to 25% for lymphomas and sarcomas.[21,22]

***Radiology.*** Second readings in radiology typically disclose discordance rates in the range of 2% to 20% for most general radiology imaging formats, although higher rates have been found in some studies.[23,24] The discordance rate in practice seems to be <5% in most cases.[25,26]

Mammography has attracted the most attention in regard to diagnostic error in radiology. There is substantial variability from one radiologist to another in the ability to accurately detect breast cancer, and it is estimated that 10% to 30% of breast cancers are missed on mammography.[27,28] A recent study of breast cancer found that the diagnosis was inappropriately delayed in 9%, and a third of these reflected misreading of the mammogram.[29] In addition to missing cancer known to be present, mammographers can be overly aggressive in reading studies, frequently recommending biopsies for what turn out to be benign lesions. Given the differences regarding insurance coverage and the medical malpractice systems between the United States and the United Kingdom, it is not surprising that women in the United States are twice as likely as women in the United Kingdom to have a negative biopsy.[30]

**Studies of Specific Conditions.** Table 1 is a sampling of studies[18,27,31–46] that have measured the rate of diagnostic error in specific conditions. An unsettling consistency emerges: the frequency of diagnostic error is disappointingly high. This is true for both relatively benign conditions and disorders where rapid and accurate diagnosis is essential, such as myocardial infarction, pulmonary embolism, and dissecting or ruptured aortic aneurysms.

**Table 1** Sampling of Diagnostic Error Rates in Specific Conditions

| Study | Conditions | Findings |
|---|---|---|
| Shojania et al (2002)[32] | Pulmonary TB | Review of autopsy studies that have specifically focused on the diagnosis of pulmonary TB; ~50% of these diagnoses were not suspected antemortem |
| Pidenda et al (2001)[33] | Pulmonary embolism | Review of fatal embolism over a 5-yr period at a single institution. Of 67 patients who died of pulmonary embolism, the diagnosis was not suspected clinically in 37 (55%) |
| Lederle et al (1994),[34] von Kodolitsch et al (2000)[35] | Ruptured aortic aneurysm | Review of all cases at a single medical center over a 7-yr period. Of 23 cases involving abdominal aneurysms, diagnosis of ruptured aneurysm was initially missed in 14 (61%); in patients presenting with chest pain, diagnosis of dissecting aneurysm of the proximal aorta was missed in 35% of cases |
| Edlow (2005)[36] | Subarachnoid hemorrhage | Updated review of published studies on subarachnoid hemorrhage: ~30% are misdiagnosed on initial evaluation |
| Burton et al (1998)[37] | Cancer detection | Autopsy study at a single hospital: of the 250 malignant neoplasms found at autopsy, 111 were either misdiagnosed or undiagnosed, and in 57 of the cases the cause of death was judged to be related to the cancer |
| Beam et al (1996)[27] | Breast cancer | 50 accredited centers agreed to review mammograms of 79 women, 45 of whom had breast cancer; the cancer would have been missed in 21% |
| McGinnis et al (2002)[18] | Melanoma | Second review of 5,136 biopsy samples; diagnosis changed in 11% (1.1% from benign to malignant, 1.2% from malignant to benign, and 8% had a change in tumor grade) |
| Perlis (2005)[38] | Bipolar disorder | The initial diagnosis was wrong in 69% of patients with bipolar disorder and delays in establishing the correct diagnosis were common |
| Graff et al (2000)[39] | Appendicitis | Retrospective study at 12 hospitals of patients with abdominal pain and operations for appendicitis. Of 1,026 patients who had surgery, there was no appendicitis in 110 (10.5%); of 916 patients with a final diagnosis of appendicitis, the diagnosis was missed or wrong in 170 (18.6%) |
| Raab et al (2005)[40] | Cancer pathology | The frequency of errors in diagnosing cancer was measured at 4 hospitals over a 1-yr period. The error rate of pathologic diagnosis was 2%–9% for gynecology cases and 5%–12% for nongynecology cases; errors represented sampling deficiencies, preparation problems, and mistakes in histologic interpretation |
| Buchweitz et al (2005)[41] | Endometriosis | Digital videotapes of laparoscopies were shown to 108 gynecologic surgeons; the interobserver agreement regarding the number of lesions was low (18%) |
| Gorter et al (2002)[42] | Psoriatic arthritis | 1 of 2 SPs with psoriatic arthritis visited 23 rheumatologists; the diagnosis was missed or wrong in 9 visits (39%) |
| Bogun et al (2004)[43] | Atrial fibrillation | Review of automated ECG interpretations read as showing atrial fibrillation; 35% of the patients were misdiagnosed by the machine, and the error was detected by the reviewing clinician only 76% of the time |
| Arnon et al (2006)[44] | Infant botulism | Study of 129 infants in California suspected of having botulism during a 5-yr period; only 50% of the cases were suspected at the time of admission |
| Edelman (2002)[45] | Diabetes mellitus | Retrospective review of 1,426 patients with laboratory evidence of diabetes mellitus (glucose >200 mg/dL* or hemoglobin $A_{1c}$ >7%); there was no mention of diabetes in the medical record of 18% of patients |
| Russell et al (1988)[46] | Chest x-rays in the ED | One third of x-rays were incorrectly interpreted by the ED staff compared with the final readings by radiologists |

ECG = electrocardiograph; ED = emergency department; SP = standardized patient; TB = tuberculosis.

*1 mg/dL = 0.05551 mmol/L.

Adapted from *Advances in Patient Safety: From Research to Implementation.*[31]

**Autopsy Studies.** The autopsy has been described as "the most powerful tool in the history of medicine"[47] and the "gold standard" for detecting diagnostic errors. Richard Cabot correlated case records with autopsy findings in several thousand patients at Massachusetts General Hospital, concluding in 1912 that the clinical diagnosis was wrong 40% of the time.[48,49] Similar discrepancies between clinical and autopsy diagnoses were found in a more recent study of geriatric patients in the Netherlands.[50] On average, 10% of autopsies revealed that the clinical diagnosis was wrong, and 25% revealed a new problem that had not been suspected clinically. Although a fraction of these discrepancies reflected incidental findings of no clinical significance, major unexpected discrepancies that potentially could have changed the outcome were found in approximately 10% of all autopsies.[32,51]

Shojania and colleagues[32] point out that autopsy studies only provide the error rate in patients who die. Because the diagnostic error rate is almost certainly lower among patients with the condition who are still alive, error rates measured solely from autopsy data may be distorted. That is, clinicians are attempting to make the diagnosis among living patients before death, so the more relevant statistic in this setting is the sensitivity of clinical diagnosis. For example, whereas autopsy studies suggest that fatal pulmonary embolism is misdiagnosed approximately 55% of the time (see Table 1), the misdiagnosis rate for all cases of pulmonary embolism is only 4%. Shojania and associates[32] argue that a large discrepancy also exists regarding the misdiagnosis rate for myocardial infarction: although autopsy data suggest roughly 20% of these events are missed, data from the clinical setting (patients presenting with chest pain or other relevant symptoms) indicate that only 2% to 4% are missed.

**Studies Using Standardized Cases.** One method of testing diagnostic accuracy is to control for variations in case presentation by using standardized cases that can enable comparisons of performance across physicians. One such approach is to incorporate what are termed *standardized patients* (SPs). Usually, SPs are lay individuals trained to portray a specific case or are individuals with certain clinical conditions trained to be study subjects.[52,53] Diagnostic errors are inevitably detected when physicians are tested with SPs or standardized case scenarios.[42,54] For example, when asked to evaluate SPs with common conditions in a clinic setting, internists missed the correct diagnosis 13% of the time.[55] Other studies using different types of standardized cases have found that not only is there variation between providers who analyze the same case[27,56] but that physicians can even disagree with themselves when presented again with a case they have previously diagnosed.[57]

## What Percentage of Adverse Events is Attributable to Diagnostic Errors and What Percentage of Diagnostic Errors Leads to Adverse Events?

Data from large-scale, retrospective, chart-review studies of adverse events have shown a high percentage of diagnostic errors. In the Harvard Medical Practice Study of 30,195 hospital records, diagnostic errors accounted for 17% of adverse events.[58,59] A more recent follow-up study of 15,000 records from Colorado and Utah reported that diagnostic errors contributed to 6.9% of the adverse events.[60] Using the same methodology, the Canadian Adverse Events Study found that 10.5% of adverse events were related to diagnostic procedures.[61] The Quality in Australian Health Care Study identified 2,351 adverse events related to hospitalization, of which 20% represented delays in diagnosis or treatment and 15.8% reflected failure to "synthesize/decide/act on" information.[62] A large study in New Zealand examined 6,579 inpatient medical records from admissions in 1998 and found that diagnostic errors accounted for 8% of adverse events; 11.4% of those were judged to be preventable.[63]

**Error Databases.** Although of limited use in quantifying the absolute incidence of diagnostic errors, voluntary error-reporting systems provide insight into the relative incidence of diagnostic errors compared with medication errors, treatment errors, and other major categories. Out of 805 voluntary reports of medical errors from 324 Australian physicians, there were 275 diagnostic errors (34%) submitted over a 20-month period.[64] Compared with medication and treatment errors, diagnostic errors were judged to have caused the most harm, but were the least preventable. A smaller study reported a 14% relative incidence of diagnostic errors from Australian physicians and 12% from physicians of other countries.[65] Mandatory error-reporting systems that rely on self-reporting typically yield fewer error reports than are found using other methodologies. For example, only 9 diagnostic errors were reported out of almost 1 million ambulatory visits over a 5.5-year period in a large healthcare system.[66]

Diagnostic errors are the most common adverse event reported by medical trainees.[67,68] Notably, of the 29 diagnostic errors reported voluntarily by trainees in 1 study, none of these were detected by the hospital's traditional incident-reporting mechanisms.[68]

**Malpractice Claims.** Diagnostic errors are typically the leading or the second-leading cause of malpractice claims in the United States and abroad.[69–72] Surprisingly, the vast majority of claims filed reflect a very small subset of diagnoses. For example, 93% of claims in the Australian registry reflect just 6 scenarios (failure to diagnose cancer, injuries after trauma, surgical problems, infections, heart attacks, and venous thromboembolic disease).[73] In a recent study of malpractice claims,[74] diagnostic errors were equally prevalent in successful and unsuccessful claims and represented 30% of all claims.

The percentage of diagnostic errors that leads to adverse events is the most difficult to determine, in that the prospective tracking needed for these studies is rarely done. As Schiff,[75] Redelmeier,[76] and Gandhi and colleagues[77] advocate, much better methods for tracking and follow-up of patients are needed. For some authors, diagnostic errors that do not result in serious harm are not even considered misdiagnoses.[78] This is little consolation, however, for the patients who suffer the consequences of these mistakes. The increasing adoption of electronic medical records, especially in ambulatory practices, will lead to better data for answering this question; research should be conducted to address this deficiency.

### Has the Diagnostic Error Rate Changed Over Time?

Autopsy data provide us the opportunity to see whether the rate of diagnostic errors has decreased over time, reflecting the many advances in medical imaging and diagnostic testing. Only 3 major studies have examined this question. Goldman and colleagues[79] analyzed 100 randomly selected autopsies from the years 1960, 1970, and 1980 at a single institution in Boston and found that the rate of misdiagnosis was stable over time. A more recent study in Germany used a similar approach to study autopsies over a range of 4 decades, from 1959 to 1989. Although the autopsy rate decreased over these years from 88% to 36%, the misdiagnosis rate was stable.[78]

Shojania and colleagues[80] propose that the near-constant rate of misdiagnosis found at autopsy over the years probably reflects 2 factors that offset each other: diagnostic accuracy actually has improved over time (more knowledge, better tests, more skills), but as the autopsy rate declines, there is a tendency to select only the more challenging clinical cases for autopsy, which then have a higher likelihood of diagnostic error. A longitudinal study of autopsies in Switzerland (constant 90% autopsy rate) supports that the absolute rate of diagnostic errors is, as suggested, decreasing over time.[81]

### Summary

In aggregate, studies consistently demonstrate a rate of diagnostic error that ranges from <5% in the perceptual specialties (pathology, radiology, dermatology) up to 10% to 15% in most other fields.

It should be noted that the accuracy of clinical diagnosis in practice may differ from that suggested by most studies assessing error rates. Some of the variability in the estimates of diagnostic errors described may be attributed to whether researchers first evaluated diagnostic errors (not all of which will lead to an adverse event) or adverse events (which will miss diagnostic errors that do not cause significant injury or disability). In addition, basing conclusions about the extent of misdiagnosis on the patients who died and had an autopsy, or who filed malpractice claims, or even who had a serious disease leads to overestimates of the extent of errors, because such samples are not representative of the vast majority of patients seen by most clinicians. On the other hand, given the fragmentation of care in the outpatient setting, the difficulty of tracking patients, and the amount of time it often takes for a clear picture of the disease to emerge, these data may actually underestimate the extent of error, especially in ambulatory settings.[82] Although the exact frequency may be difficult to determine precisely, it is clear that an extensive and ever-growing literature confirms that diagnostic errors exist at nontrivial and sometimes alarming rates. These studies span every specialty and virtually every dimension of both inpatient and outpatient care.

## PHYSICIAN OVERCONFIDENCE

> *". . . what discourages autopsies is medicine's twenty-first century, tall-in-the-saddle confidence."*
> *"When someone dies, we already know why. We don't need an autopsy to find out. Or so I thought."*
>
> —Atul Gawande[83]
>
> *"He who knows best knows how little he knows."*
>
> —attributed to Thomas Jefferson[84]
>
> *"Doctors think a lot of patients are cured who have simply quit in disgust."*
>
> —attributed to Don Herold[85]

As Kirch and Schafii[78] note, autopsies not only document the presence of diagnostic errors, they also provide an opportunity to learn from one's errors (*errando discimus*) if one takes advantage of the information. The rate of autopsy in the United States is not measured any more, but is widely assumed to be significantly <10%. To the extent that this important feedback mechanism is no longer a realistic option, clinicians have an increasingly distorted view of their own error rates. In addition to the lack of autopsies, as the above quote by Gawande indicates, physician overconfidence may prevent them from taking advantage of these important lessons. In this section, we review studies related to physician overconfidence and explore the possibility that this is a major factor contributing to diagnostic error.[86] Overconfidence may have both attitudinal as well as cognitive components and should be distinguished from complacency.

There are several reasons for separating the various aspects of overconfidence and complacency: (1) Some areas have undergone more research than others. (2) The strategies for addressing these 2 qualities may be different. (3) Some aspects are more amenable to being addressed than others. (4) Some may be a more frequent cause of misdiagnoses than others.

### Attitudinal Aspects of Overconfidence

This aspect (i.e., "I know all I need to know") is reflected within the more pervasive attitude of *arrogance*, an outlook that expresses disinterest in any decision support or feedback, regardless of the specific situation.

Comments like those quoted at the beginning of this section reflect the perception that physicians are arrogant and pervasively overconfident about their abilities; however, the data on this point are mostly indirect. For example, the evidence discussed above—that autopsies are on the decline despite their providing useful data—inferentially provides support for the conclusion that physicians do not think they need diagnostic assistance. Substantially more data are available on a similar line of evidence, namely, the general tendency on the part of physicians to disregard, or fail to use, decision-support resources.

**Knowledge-Seeking Behavior.** Research shows that physicians admit to having many questions that could be important at the point of care, but which they do not pursue.[87–89] Even when information resources are automated and easily accessible at the point of care with a computer, Rosenbloom and colleagues[90] found that a tiny fraction of the resources were actually used. Although the method of accessing resources affected the degree to which they were used, even when an indication flashed on the screen that relevant information was available, physicians rarely reviewed it.

**Response to Guidelines and Decision-Support Tools.** A second area related to the attitudinal aspect is research on physician response to clinical guidelines and to output from computerized decision-support systems, often in the form of guidelines, alerts, and reminders. A comprehensive review of medical practice in the United States found that the care provided deviated from recommended best practices half of the time.[91] For many conditions, consensus exists on the best treatments and the recommended goals; nevertheless, these national clinical guidelines have a high rate of noncompliance.[92,93] The treatment of high cholesterol is a good example: although 95% of physicians were aware of lipid treatment guidelines from a recent study, they followed these guidelines only 18% of the time.[94] Decision-support tools have the potential to improve care and decrease variations in care delivery, but, unfortunately, clinicians disregard them, even in areas where care is known to be suboptimal and the support tool is well integrated into their workflow.[95–99]

In part, this disregard reflects the inherent belief on the part of many physicians that their practice conforms to consensus recommendations, when in fact it does not. For example, Steinman and colleagues[100] were unable to find a significant correlation between perceived and actual adherence to hypertension treatment guidelines in a large group of primary care physicians.

Similarly, because treatment guidelines are frequently dependent on accurate diagnoses, if the clinician does not recognize the diagnosis, the guideline may not be invoked. For instance, Tierney and associates[101] implemented computer-based guidelines for asthma that did not work successfully, in part because physicians did not consider certain cases to be asthma even though they met identified clinical criteria for the condition.

Timmermans and Mauck[102] suggest that the high rate of noncompliance with clinical guidelines relates to the sociology of what it means to be a professional. Being a professional connotes possessing expert knowledge in an area and functioning relatively autonomously. In a similar vein, Tanenbaum[103] worries that evidence-based medicine will decrease the "professionalism" of the physician. van der Sijs and colleagues[104] suggest that the frequent overriding of computerized alerts may have a positive side in that it shows clinicians are not becoming overly dependent on an imperfect system. Although these authors focus on the positive side to professionalism, the converse, a pervasive attitude of overconfidence, is certainly a possible explanation for the frequent overrides. At the very least, as Katz[105] noted many years ago, the discomfort in admitting uncertainty to patients that many physicians feel can mask inherent uncertainties in clinical practice even to the physicians themselves. Physicians do not tolerate uncertainty well, nor do their patients.

## Cognitive Aspects of Overconfidence

The cognitive aspect (i.e., "not knowing what you don't know") is situation specific, that is, in a particular instance, the clinician thinks he/she has the correct diagnosis, but is wrong. Rarely, the reason for not knowing may be lack of knowledge per se, such as seeing a patient with a disease that the physician has never encountered before. More commonly, cognitive errors reflect problems gathering data, such as failing to elicit complete and accurate information from the patient; failure to recognize the significance of data, such as misinterpreting test results; or most commonly, failure to synthesize or "put it all together."[106] This typically includes a breakdown in clinical reasoning, including using faulty heuristics or "cognitive dispositions to respond," as described by Croskerry.[107] In general, the cognitive component also includes a failure of metacognition (the willingness and ability to reflect on one's own thinking processes and to critically examine one's own assumptions, beliefs, and conclusions).

**Direct Evidence of Overconfidence.** A direct approach to studying overconfidence is to simply ask physicians how confident they are in their diagnoses. Studies examining the cognitive aspects of overconfidence generally have examined physicians' expressed confidence in specific diagnoses, usually in controlled "laboratory" settings rather than studies in actual practice settings. For instance, Friedman and colleagues[108] used case scenarios to examine the accuracy of physicians', residents', and medical students' actual diagnoses compared with how confident they were that their diagnoses were correct. The researchers found that residents had the greatest mismatch. That is, medical students were both least accurate and least confident, whereas attending physicians were the most accurate and highly confident. Residents, on the other hand, were more confident about the correctness of their diagnoses, but they were less accurate than the attending physicians.

Berner and colleagues,[99] while not directly assessing confidence, found that residents often stayed wedded to an incorrect diagnosis even when a diagnostic decision support system suggested the correct diagnosis. Similarly, experienced dermatologists were confident in diagnosing melanoma in >50% of test cases, but were wrong in 30% of these decisions.[109] In test settings, physicians are also overconfident in treatment decisions.[110] These studies were done with simulated clinical cases in a formal research setting and, although suggestive, it is not clear that the results would be the same with cases seen in actual practice.

Concrete and definite evidence of overconfidence in medical practice has been demonstrated at least twice, using autopsy findings as the gold standard. Podbregar and colleagues[111] studied 126 patients who died in the ICU and underwent autopsy. Physicians were asked to provide the clinical diagnosis and also their level of uncertainty: level 1 represented complete certainty, level 2 indicated minor uncertainty, and level 3 designated major uncertainty. The rates at which the autopsy showed significant discrepancies between the clinical and postmortem diagnosis were essentially identical in all 3 of these groups. Specifically, clinicians who were "completely certain" of the diagnosis antemorten were wrong 40% of the time.[111] Similar findings were reported by Landefeld and coworkers[112]: the level of physician confidence showed no correlation with their ability to predict the accuracy of their clinical diagnosis. Additional direct evidence of overconfidence has been demonstrated in studies of radiologists given sets of "unknown" films to classify as normal or abnormal. Potchen[113] found that diagnostic accuracy varied among a cohort of 95 board-certified radiologists: The top 20 had an aggregate accuracy rate of 95%, compared with 75% for the bottom 20. Yet, the confidence level of the worst performers was actually higher than that of the top performers.

**Causes of Cognitive Error.** Retrospective studies of the accuracy of diagnoses in actual practice, as well as the autopsy and other studies described previously,[77,106,114,115] have attempted to determine reasons for misdiagnosis. Most of the cognitive errors in diagnosis occur during the "synthesis" step, as the physician integrates his/her medical knowledge with the patient's history and findings.[106] This process is largely subconscious and automatic.

***Heuristics.*** Research on these automatic responses has revealed a wide variety of *heuristics* (subconscious rules of thumb) that clinicians use to solve diagnostic puzzles.[116] Croskerry[107] calls these responses our "cognitive predispositions to respond." These heuristics are powerful clinical tools that allow problems to be solved quickly and, typically, correctly. For example, a clinician seeing a weekend gardener with linear streaks of intensely itchy vesicles on the legs easily diagnoses the patient as having a contact sensitivity to poison ivy using the *availability heuristic*. He or she has seen many such reactions because this is a common problem, and it is the first thing to come to mind. The *representativeness heuristic* would be used to diagnose a patient presenting with chest pain if the pain radiates to the back, varies with posture, and is associated with a cardiac friction rub. This patient has pericarditis, an extremely uncommon reason for chest pain, but a condition with a characteristic clinical presentation.

Unfortunately, the unconscious use of heuristics can also predispose to diagnostic errors. If a problem is solved using the availability heuristic, for example, it is unlikely that the clinician considers a comprehensive differential diagnosis, because the diagnosis is so immediately obvious, or so it appears. Similarly, using the representativeness heuristic predisposes to base rate errors. That is, by just matching the patient's clinical presentation to the prototypical case, the clinician may not adequately take into account that other diseases may be much more common and may sometimes present similarly.

Additional cognitive errors are described below. Of these, premature closure and the context errors are the most common causes of cognitive error in internal medicine.[86]

***Premature Closure.*** Premature closure is narrowing the choice of diagnostic hypotheses too early in the process, such that the correct diagnosis is never seriously considered.[117–119] This is the medical equivalent of Herbert Simon's concept of "satisficing."[120] Once our minds find an adequate solution to whatever problem we are facing, we tend to stop thinking of additional, potentially better solutions.

***Confirmation Bias and Related Biases.*** These biases reflect the tendency to seek out data that confirm one's original idea rather than to seek out disconfirming data.[115]

***Context Errors.*** Very early in clinical problem solving, healthcare practitioners start to characterize a problem in terms of the organ system involved, or the type of abnormality that might be responsible. For example, in the instance of a patient with new shortness of breath and a past history of cardiac problems, many clinicians quickly jump to a diagnosis of congestive heart failure, without consideration of other causes of the shortness of breath. Similarly, a patient with abdominal pain is likely to be diagnosed as having a gastrointestinal problem, although sometimes organs in the chest can present in this fashion. In these situations, clinicians are biased by the history, a previously established diagnosis, or other factors, and the case is formulated in the wrong context.

**Clinical Cognition.** Relevant research has been conducted on how physicians make diagnoses in the first place. Early work by Elstein and associates,[121] and Barrows and colleagues[122–124] showed that when faced with what is perceived as a difficult diagnostic problem, physicians gather some initial data and very quickly often within seconds, develop diagnostic hypotheses. They then gather more data to evaluate these hypotheses and finally reach a diagnostic conclusion. This approach has been referred to as a hypothetico-deductive mode of diagnostic reasoning and is similar to the traditional descriptions of the scientific method.[121] It is during this evaluation process that the problems of confirmation bias and premature closure are likely to occur.

Although hypothetico-deductive models may be followed for situations perceived as diagnostic challenges, there is also evidence that as physicians gain experience and expertise, most problems are solved by some sort of pattern-recognition process, either by recalling prior similar cases, attending to prototypical features, or other similar strategies.[125–129] As Eva and Norman[130] and Klein[128] have emphasized, most of the time this pattern recognition serves the clinician well. However, it is during the times when it does not work, whether because of lack of knowledge or because of the inherent shortcomings of heuristic problem solving, that overconfidence may occur.

There is substantial evidence that overconfidence— that is, miscalibration of one's own sense of accuracy and actual accuracy—is ubiquitous and simply part of human nature. Miscalibration can be easily demonstrated in experimental settings, almost always in the direction of overconfidence.[84,131–133] A striking example derives from surveys of academic professionals, 94% of whom rate themselves in the top half of their profession.[134] Similarly, only 1% of drivers rate their skills below that of the average driver.[135] Although some attribute the results to statistical artifacts, and the degree of overconfidence can vary with the task, the inability of humans to accurately judge what they know (in terms of accuracy of judgment or even thinking that they know or do not know something) is found in many areas and in many types of tasks.

Most of the research that has examined expert decision making in natural environments, however, has concluded that rapid and accurate pattern recognition is characteristic of experts. Klein,[128] Gladwell,[127] and others have examined how experts in fields other than medicine diagnose a situation and find that they routinely rapidly and accurately assess the situation and often cannot even describe how they do it. Klein[128] refers to this process as "recognition primed" decision making, referring to the extensive experience of the expert with previous similar cases. Gigerenzer and Goldstein[136] similarly support the concept that most real-world decisions are made using automatic skills, with "fast and frugal" heuristics that lead to the correct decisions with surprising frequency.

Again, when experts recognize that the pattern is incorrect they may revert back to a hypothesis testing mode or may run through alternative scripts of the situation. Expertise is characterized by the ability to recognize when one's initial impression is wrong and to having back-up strategies readily available when the initial strategy does not work.

Hamm[137] has suggested that what is known as the cognitive continuum theory can explain some of the contradictions as to whether experts follow a hypothetico-deductive or a pattern-recognition approach. The cognitive continuum theory suggests that clinical judgment can appropriately range from more intuitive to more analytic, depending on the task. Intuitive judgment, as Hamm conceives it, is not some vague sense of intuition, but is really the rapid pattern

acteristic of experts in many situations. Although intuitive judgment may be most appropriate in the uncertain, fast-paced field environment where Klein observed his subjects, other strategies might best suit the laboratory environment that others use to study decision making. In addition, forcing research subjects to verbally explain their strategies, as done in most experimental studies of physician problem solving, may lead to the hypothetico-deductive description. In contrast, Klein,[128] who studied experts in field situations, found his subjects had a very difficult time articulating their strategies.

Even if we accept that a pattern-recognition strategy is appropriate under some circumstances and for certain types of tasks, we are still left with the question as to whether overconfidence is in fact a significant problem. Gigerenzer[138] (like Klein) feels that most of the formal studies of cognition leading to the conclusion of overconfidence use tasks that are not representative of decision making in the real world, either in content or in difficulty. As an example, to study diagnostic problem solving, most researchers of necessity use "diagnostically challenging cases,"[139] which are clearly not typical of the range of cases seen in clinical practice. The zebra adage (i.e., when you hear hoofbeats think of horses, not zebras) may for the most part be adaptive in the clinicians' natural environment, where zebras are much rarer than horses. However, in experimental studies of clinician diagnostic decision making, the reverse is true. The challenges of studying clinicians' diagnostic accuracy in the natural environment are compounded by the fact that most initial diagnoses are made in ambulatory settings, which are notoriously difficult to assess.[82]

## Complacency Aspect of Overconfidence

Complacency (i.e., "nobody's perfect") reflects a combination of underestimation of the amount of error, tolerance of error, and the belief that errors are inevitable. Complacency may show up as thinking that misdiagnoses are more infrequent than they actually are, that the problem exists but not in the physician's own practice, that other problems are more important to address, or that nothing can be done to minimize diagnostic errors.

Given the overwhelming evidence that diagnostic error exists at nontrivial rates, one might assume that physicians would appreciate that such error is a serious problem. Yet this is not the case. In 1 study, family physicians asked to recall memorable errors were able to recall very few.[140]

However, 60% of those recalled were diagnostic errors. When giving talks to groups of physicians on diagnostic errors, Dr. Graber (coauthor of this article) frequently asks whether they have made a diagnostic error in the past year. Typically, only 1% admit to having made a diagnostic error. The concept that they, personally, could err at a significant rate is inconceivable to most physicians.

While arguing that clinicians grossly underestimate their own error rates, we accept that they are generally aware of the problem of medical error, especially in the context of medical malpractice. Indeed, 93% of physicians in formal surveys reported that they practice "defensive medicine," including ordering unnecessary lab tests, imaging studies, and consultations.[141] The cost of defensive medicine is estimated to consume 5% to 9% of healthcare expenditures in the United States.[142] We conclude that physicians acknowledge the possibility of error, but believe that mistakes are made by others.

The remarkable discrepancy between the known prevalence of error and physician perception of their own error rate has not been formally quantified and is only indirectly discussed in the medical literature, but lies at the crux of the diagnostic error puzzle, and explains in part why so little attention has been devoted to this problem. Physicians tend to be overconfident of their diagnoses and are largely unaware of this tendency at any conscious level. This may reflect either inherent or learned behaviors of self-deception. Self-deception is thought to be an everyday occurrence, serving to emphasize to others our positive qualities and minimize our negative ones.[143] From the physician's perspective, such self-deception can have positive effects. For example, it can help foster the patient's perception of the physician as an all-knowing healer, thus promoting trust, adherence to the physician's advice, and an effective patient-physician relationship.

Other evidence for complacency can be seen in data from the review by van der Sijs and colleagues.[104] The authors cite several studies that examined the outcomes of the overrides of automated alerts, reminders, and guidelines. In many cases, the overrides were considered clinically justified, and when they were not, there were very few ($\leq$3%) adverse events as a result. While it may be argued that even those few adverse events could have been averted, such contentions may not be convincing to a clinician who can point to adverse events that occur even with adherence to guidelines or alerts. Both types of adverse events may appear to be unavoidable and thus reinforce the physician's complacency.

Gigerenzer,[138] like Eva and Norman[130] and Klein,[128] suggests that many strategies used in diagnostic decision making are adaptive and work well most of the time. For instance, physicians are likely to use data on patients' health outcome as a basis for judging their own diagnostic acumen. That is, the physician is unconsciously evaluating the number of clinical encounters in which patients improve compared with the overall number of visits in a given period of time, or more likely, over years of practice. The denominator that the clinician uses is clearly not the number of adverse events, which some studies of diagnostic errors have used. Nor is it a selected sample of challenging cases, as others have cited. Because most visits are not diagnostically challenging, the physician not only is going to diagnose most of these cases appropriately but he/she also is likely to get accurate feedback to that effect, in that most patients (1) do not wind up in the hospital, (2) appear to be satisfied when next seen, or (3) do not return for the particular complaint because they are cured or treated appropriately.

Causes of inadequate feedback include patients leaving the practice, getting better despite the wrong diagnosis, or returning when symptoms are more pronounced and thus eventually getting diagnosed correctly. Because immediate feedback is not even expected, feedback that is delayed or absent may not be recognized for what it is, and the perception that "misdiagnosis is not a big problem" remains unchallenged. That is, in the absence of information that the diagnosis is wrong, it is assumed to be correct ("no news is good news"). This phenomenom is illustrated in epigraph above from Herold, "Doctors think a lot of patients are cured who have simply quit in disgust."[85] The perception that misdiagnosis is not a major problem, while not necessarily correct, may indeed reflect arrogance, "tall in the saddle confidence,"[83] or "omniscience."[144] Alternatively, it may simply reflect that over all the patient encounters a physician has, the number of diagnostic errors *of which he or she is aware* is very low.

Thus, despite the evidence that misdiagnoses do occur more frequently than often presumed by clinicians, and despite the fact that recognizing that they do occur is the first step to correcting the problem, the assumption that misdiagnoses are made only a very small percentage of the time can be seen as a rational conclusion given the current healthcare environment where feedback is limited and only selective outcome data are available for physicians to accurately calibrate the extent of their own misdiagnoses.

## Summary

Pulling together the research described above, we can see why there may be complacency and why it is difficult to address. First, physicians generate hypotheses almost immediately upon hearing a patient's initial symptom presentation and in many cases these hypotheses suggest a familiar pattern. Second, even if more exploration is needed, the most likely information sought is that which confirms the initial hypothesis; often, a decision is reached without full exploration of a large number of other possibilities. In the great majority of cases, this approach leads to the correct diagnosis and a positive outcome. The patient's diagnosis is made quickly and correctly, treatment is initiated, and both the patient and physician feel better. This explains why this approach is used, and why it is so difficult to change. In addition, in many of the cases where the diagnosis is incorrect, the physician never knows it. If the diagnostic process routinely led to errors that the physician recognized, they could get corrected. Additionally, the physician might be humbled by the frequent oversights and become inclined to adopt a more deliberate, contemplative approach or develop strategies to better identify and prevent the misdiagnoses.

## STRATEGIES TO IMPROVE THE ACCURACY OF DIAGNOSTIC DECISION MAKING

> *"Ignorance more frequently begets confidence than does knowledge."*
>
> —Charles Darwin, 1871[145]

We believe that strategies to reduce misdiagnoses should focus on physician calibration, i.e., improving the match between the physician's self-assessment of errors and actual errors. Klein[128] has shown that experts use their intuition on a routine basis, but rethink their strategies when that does not work. Physicians also rethink their diagnoses when it is obvious that they are wrong. In fact, it is in these situations that diagnostic decision-support tools are most likely to be used.[146]

The challenge becomes how to increase physicians' awareness of the possibility of error. In fact, it could be argued that their awareness needs to be increased for a select type of case: that in which the healthcare provider thinks he/she is correct and does not receive any timely feedback to the contrary, but where he/she is, in fact, mistaken. Typically, most of the clinician's cases are diagnosed correctly; these do not pose a problem. For the few cases where the clinician is consciously puzzled about the diagnosis, it is likely that an extended workup, consultation, and research into possible diagnoses occurs. It is for the cases that fall between these types, where miscalibration is present but unrecognized, that we need to focus on strategies for increasing physician awareness and correction.

If overconfidence, or more specifically, miscalibration, is a problem, what is the solution? We examine 2 broad categories of solutions: strategies that focus on the individual and system approaches directed at the healthcare environment in which diagnosis takes place. The individual approaches assume that the physician's cognition needs improvement and focus on making the clinician smarter, a better thinker, less subject to biases, and more cognizant of what he or she knows and does not know. System approaches assume that the individual physician's cognition is adequate for the diagnostic and metacognitive tasks, but that he/she needs more, and better, data to improve diagnostic accuracy. Thus, the system approaches focus on changing the healthcare environment so that the data on the patients, the potential diagnoses, and any additional information are more accurate and accessible. These 2 approaches are not mutually exclusive and the major aim of both is to improve the physician's calibration between his/her perception of the case and the actual case. Theorectically, if improved calibration occurs, overconfidence should decrease, including the attitudinal components of arrogance and complacency.

In the discussion about individually focused solutions, we review the effectiveness of clinical education and practice, development of metacognitive skills, and training in reflective practice. In the section on systems-focused solutions, we examine the effectiveness of providing performance feedback, the related area of improving follow-up of patients and their health outcomes, and using automation—such as providing general knowledge resources at the point of care and specific diagnostic decision-support programs.

### Strategies that Focus on the Individual

**Education, Training and Practice.** By definition, experts are smarter, e.g., more knowledgeable than novices. A fascinating (albeit frightening) observation is the general tendency of novices to overrate their skills.[84,108,132] Exactly the same tendency is seen in testing of medical trainees in regard to skills such as communicating with patients.[147] In a typical experiment a cohort with varying degrees of expertise are asked to undertake a skilled task. At the completion of the task, the test subjects are asked to grade their own performance. When their self-rated scores are compared with the scores assigned by experts, the individuals with the lowest skill levels predictably overestimate their performance.

Data from a study conducted by Friedman and colleagues[108] showed similar results: residents in training performed worse than faculty physicians, but were more confident in the correctness of their diagnoses. A systematic review of studies assessing the accuracy of physicians' self-assessment of knowledge compared with an external measure of competence showed very little correlation between self-assessment and objective data.[148] The authors also found that those physicians who were least expert tended to be most overconfident in their self-assessments.

These observations suggest a possible solution to overconfidence: make physicians more expert. The expert is better calibrated (i.e. better assesses his/her own accuracy), and excels at distinguishing cases that are easily diagnosed from those that require more deliberation. In addition to their enhanced ability to make this distinction, experts are likely to make the correct diagnosis more often in both recognized as well as unrecognized cases. Moreover, experts carry out these functions automatically, more efficiently, and with less resource consumption than nonexperts.[127,128]

The question, of course, is how to develop that expertise. Presumably, thorough medical training and continuing education for physicians would be useful; however, data show that the effects on actual practice of many continuing education programs are minimal.[149–151] Another approach is to advocate the development of expertise in a narrow domain. This strategy has implications for both individual clinicians and healthcare systems. At the level of the individual clinician, the mandate to become a true expert would drive more trainees into subspecialty training and emphasize development of a comprehensive knowledge base.

Another mechanism for gaining knowledge is to gain more extensive practice and experience with actual clinical cases. Both Bordage[152] and Norman[151,153] champion this approach, arguing that "practice is the best predictor of performance." Having a large repertoire of mentally stored exemplars is also the key requirement for Gigerenzer's "fast and frugal"[136,138] and Klein's[128] "recognition-primed" decision making. Extensive practice with simulated cases may supplement, although not supplant, experience with real ones. The key requirements in regard to clinical practice are extensive, i.e., necessitating more than just a few cases and occasional feedback.

**Metacognitive Training and Reflective Practice.** In addition to strategies that aim to increase the overall level of clinicians' knowledge, other educational approaches focus on increasing physicians' self-awareness so that they can recognize when additional information is needed or the wrong diagnostic path is taken. One such approach is to increase what has been called "situational awareness,'" the lack of which has been found to lie behind errors in aviation.[154] Singh and colleagues[154] advocate this strategy; their definition of types of situational awareness is similar to what others have called metacognitive skills. Croskerry[115,155] and Hall[156] champion the idea that metacognitive training can reduce diagnostic errors, especially those involving subconscious processing. The logic behind this approach is appealing: Because much of intuitive medical decision making involves the use of cognitive dispositions to respond, the assumption is if trainees or clinicians were educated about the inherent biases involved in the use of these strategies, they would be less susceptible to decision errors.

Croskerry[157] has outlined the use of what he refers to as "cognitive forcing strategies" to counteract the tendency to cognitive error. These would orient clinicians to the general concepts of metacognition (a universal forcing strategy), familiarize them with the various heuristics they use intuitively and their associated biases (generic forcing strategies), and train them to recognize any specific pitfalls that apply to the types of patients they see most commonly (specific forcing strategies).

Another noteworthy approach developed by the military, which suggests focusing on a comprehensive conscious view of the proposed diagnosis and how this was derived, is the technique of *prospective hindsight*.[158] Once the initial diagnosis is made, the clinician figuratively gazes into a crystal ball to see the future, sees that the initial diagnosis is not correct, and is thus forced to consider what else it could it be. A related technique, which is taught in every medical school, is to construct a comprehensive differential diagnosis on each case before planning an appropriate workup. Although students and residents excel at this exercise, they rarely use it outside the classroom or teaching rounds. As we discussed earlier, with more experience, clinicians begin to use a pattern-recognition approach rather than an exhaustive differential diagnosis. Other examples of cognitive forcing strategies include advice to always "consider the opposite," or ask "what diagnosis can I not afford to miss?"[76] Evidence that metacognitive training can decrease the rate of diagnostic errors is not yet available, although preliminary results are encouraging.[156]

*Reflective practice* is an approach defined as the ability of physicians to critically consider their own reasoning and decisions during professional activities.[159] This incorporates the principles of metacognition and 4 additional attributes: (1) the tendency to search for alternative hypotheses when considering a complex, unfamiliar problem; (2) the ability to explore the consequences of these alternatives; (3) a willingness to test any related predictions against the known facts; and (4) openness toward reflection that would allow for better toleration of uncertainty.[160] Experimental studies show that reflective practice enhances diagnostic accuracy in complex situations.[161] However, even advocates of this approach recognize that it is an untested assumption in terms of whether lessons learned in educational settings can transfer to the practice setting.[162]

## System Approaches

One could argue that effectively incorporating the education and training described above would require system-level change. For instance, at the level of healthcare systems, in addition to the development of required training and education, a concerted effort to increase the level of expertise of the individual would require changes in staffing policies and access to specialists.

If they are designed to teach the clinician, or at least function as an adjunct to the clinician's expertise, some decision-support tools also serve as systems-level interventions that have the potential to increase the total expertise available. If used correctly, these products are designed to allow the less expert clinician to function like a more expert clinician. Computer- or web-based information sources also may serve this function. These resources may not be very different from traditional knowledge resources (e.g., medical books and journals), but by making them more accessible at the point of care they are likely to be used more frequently (assuming the clinician has the metacognitive skills to recognize when they are needed).

The systems approaches described below are based on the assumption that both the knowledge and metacognitive skills of the healthcare provider are generally adequate. These approaches focus on providing better and more accurate information to the clinician primarily to improve calibration. James Reason's ideas on systems approaches for reducing medical errors have formed the background of the patient safety movement, although they have not been applied specifically to diagnostic errors.[163] Nolan[164] advocates 3 main strategies based on a systems approach: prevention, making error visible, and mitigating the effects of error. Most of the cognitive strategies described above fall into the category of prevention.

The systems approaches described below fall chiefly into the latter two of Nolan's strategies. One approach is to provide expert consultation to the physician. Usually this is done by calling in a consultant or seeking a second opinion.

A second approach is to use automated methods to provide diagnostic suggestions. Usually a diagnostic decision-support system is used once the error is visible (e.g., the clinician is obviously puzzled by the clinical situation). Using the system may prevent an initial misdiagnosis and may also mitigate possible sequelae.

**Computer-based Diagnostic Decision Support.** A variety of diagnostic decision-support systems were developed out of early expert system research. Berner and colleagues[139] performed a systematic evaluation of 4 of these systems; in 1994, Miller[165] described these and other systems. In a review article. Miller's overall conclusions were that while the niche systems for well-defined specific areas were clearly effective, the perceived usefulness of the more general systems such as Quick Medical Reference (QMR), DXplain, Iliad, Meditel was less certain, despite evidence that they could suggest diagnoses that even expert physicians had not considered. The title, "A Report Card on Computer-Assisted Diagnosis—The Grade Is C," of Kassirer's editorial[166] that accompanied the article by Berner and associates[139] is illustrative of an overall negative attitude toward these systems. In a subsequent study, Berner and colleagues[167] found that less experienced physicians were more likely than more experienced physicians to find QMR useful; some researchers have suggested that these systems may be more useful in educational settings.[168] Lincoln and colleagues[169–171] have shown the effectiveness of the Iliad system in educational settings. Arene and associates[172] showed that QMR was effective in improving residents' diagnoses, but then concluded that it took too much time to learn to use the system.

A similar response was found more recently in a randomized controlled trial of another decision-support system (Problem-Knowledge Couplers (PKC), Burlington, Vt).[173] Users felt that the information provided by PKC was useful, but that it took too much time to use. More disturbing was that use of the system actually increased costs, perhaps by suggesting more diagnoses to rule out. What is interesting about PKC is that in this system the patient rather than the physician enters all the data, so the complaint that the system required too much time most likely reflected physician time to review and discuss the results rather than data entry.

One of the more recent entries into the diagnostic decision-support system arena is Isabel (Isabel Healthcare, Inc., Reston, VA; Isabel Healthcare, Ltd., Haslemere, UK.) which was initially begun as a pediatric system and now is also available for use in adults.[174–178] The available studies using Isabel show that it provides diagnoses that are considered both accurate and relevant by physicians. Both Miller[179] and Berner[180] have reviewed the challenges in evaluating medical diagnostic programs. Basically, it is difficult to determine the gold standard against which the systems should be evaluated, but both investigators advocate that the criterion should be how well the clinician using the computer compares with use of only his/her own cognition.[179,180] Virtually all of the published studies have evaluated these systems only in artificial situations and many of them have been performed by the developers themselves.

The history of these systems is reflective of the overall problem we have demonstrated in other domains: despite evidence that these systems can be helpful, and despite studies showing users are satisfied with their results when they do use them, many physicians are simply reluctant to use decision-support tools in practice.[181] Meditel, QMR, and Iliad are no longer commercially available. DXplain, PKC, and Isabel are still available commercially, but although there may be data on the extent of use, there are no data on how often they are used compared with how often they could/should have been used. The study by Rosenbloom and colleagues,[90] which used a well-integrated, easy-to-access system, showed that clinicians very rarely take advantage of the available opportunities for decision support. Because diagnostic tools require the user to enter the data into the programs, it is likely that their usage would be even lower or that the data entry may be incomplete.

An additional concern is that the output of most of these decision-support programs requires subsequent mental filtering, because what is usually displayed is a (sometimes lengthy) list of diagnostic considerations. As we have discussed previously, not only does such filtering take time,[173] but the user must be able to distinguish likely from unlikely diagnoses, and data show that such recognition can be difficult.[99] Also, as Teich and colleagues[182] noted with other decision-support tools, physicians accept reminders about things they intend to do, but are less willing to accept advice that forces them to change their plans. It is likely that if physicians already have a work-up strategy in mind, or are sure of their diagnoses, they would be less willing to consult such a system. For many clinicians, these factors may make the perceived utility of these systems not worth the cost and effort to use them. That does not mean that they are not potentially useful, but the limited interest in them has made several commercial ventures unsustainable.

In summary, the data on diagnostic decision-support systems in reducing diagnostic errors shows that they can provide what are perceived as useful diagnostic suggestions. Every commercial system also has what amounts to testimonials about its usefulness in real life—stories of how the system helped the clinician recognize a rare disease[146] — but to date their use in actual clinical situations has been limited to those times that the physician is puzzled by a diagnostic problem. Because such puzzles occur rarely, there is not enough use of the systems in real practice situations to truly evaluate their effectiveness.

**Feedback and Calibration.** A second general category of a systems approach is to design systems to provide feedback to the clinician. Overconfidence represents a mismatch between perceived and actual performance. It is a state of miscalibration that, according to existing paradigms of cognitive psychology, should be correctable by providing feedback. Feedback in general can serve to make the diagnostic error visible, and timely feedback can mitigate the harm that the initial misdiagnosis might have caused. Accurate feedback can improve the basis on which the clinicians are judging the frequency of events, which may improve calibration.

Feedback is an essential element in developing expertise. It confirms strengths and identifies weaknesses, guiding the way to improved performance. In this framework, a possible approach to reducing diagnostic error, overconfidence, and error-related complacency is to enhance feedback with the goal of improving calibration.[183]

Experiments confirm that feedback can improve performance,[184] especially if the feedback includes cognitive information (for example, why a certain diagnosis is favored) as opposed to simple feedback on whether the diagnosis was correct or not.[185,186] A recent investigation by Sieck and Arkes,[131] however, emphasizes that overconfidence is highly ingrained and often resistant to amelioration by simple feedback interventions.

The timing of feedback is important. Immediate feedback is effective, delayed feedback less so.[187] This is particularly problematic for diagnostic feedback in real clinical settings, outside of contrived experiments, because such feedback often is not available at all, much less immediately or soon after the diagnosis is made. In fact, the gold standard for feedback regarding clinical judgment is the autopsy, which of course can only provide retrospective, not real-time, diagnostic feedback.

Radiology and pathology are the only fields of medicine where feedback has been specifically considered, and in some cases adopted, as a method of improving performance and calibration.

***Radiology.*** The accuracy of radiologic diagnosis is most sharply focused in the area of mammography, where both false-positive and false-negative reports have substantial clinical impact. Of note, a recent study called attention to an interesting difference between radiologists in the United States and their counterparts in the United Kingdom: US radiologists suggested follow-up studies (more radiologic testing, biopsy, or close clinical follow-up) twice as often as UK radiologists, and US patients had twice as many normal biopsies, whereas the cancer detection rates in the 2 countries were comparable.[30] In considering the reasons for this difference in performance, the authors point out that 85% of mammographers in the United Kingdom voluntarily participate in "PERFORMS," an organized calibration process, and 90% of programs perform double readings of mammograms. In contrast, there are no organized calibration exercises in the United States and few programs require "double reads." An additional difference is the expectation for accreditation: US radiologists must read 480 mammograms annually to meet expectations of the Mammography Quality Standards Act, whereas the comparable expectation for UK mammographers is 5,000 mammograms per year.[30]

As an initial step toward performance improvement by providing organized feedback, the American College of Radiology (ACR) recently developed and launched the "RADPEER" process.[188] In this program, radiologists keep track of their agreement with any prior imaging studies they re-review while they are evaluating a current study, and the ACR provides a mechanism to track these scores. Participation is voluntary; it will be interesting to see how many programs enroll in this effort.

***Pathology.*** In response to a *Wall Street Journal* exposé on the problem of false-negative Pap smears, the US Congress enacted the Clinical Laboratory Improvement Act of 1988. This act mandated more rigorous quality measures in regard to cytopathology, including proficiency testing and mandatory reviews of negative smears.[189] Even with these measures in place, however, rescreening of randomly selected smears discloses a discordance rate in the range of 10% to 30%, although only a fraction of these discordances have major clinical impact.[190]

There are no comparable proficiency requirements for anatomic pathology, other than the voluntary "Q-Probes" and "Q-Tracks" programs offered by the College of American Pathologists (CAP). Q-Probes are highly focused reviews that examine individual aspects of diagnostic testing, including preanalytical, analytical, and postanalytical errors. The CAP has sponsored hundreds of these probes. Recent examples include evaluating the appropriateness of testing for $\beta$-natriuretic peptides, determining the rate of urine sediment examinations, and assessing the accuracy of send-out tests. Q-Tracks are monitors that "reach beyond the testing phase to evaluate the processes both within and beyond the laboratory that can impact test and patient outcomes."[191] Participating labs can track their own data and see comparisons with all other participating labs. Several monitors evaluate the accuracy of diagnosis by clinical pathologists and cytopathologists. For example, participating centers can track the frequency of discrepancies between diagnoses suggested from Pap smears compared with results obtained from biopsy or surgical specimens. However, a recent review estimated that <1% of US programs participate in these monitors.[192]

Pathology and radiology are 2 specialties that have pioneered the development of computerized second opinions. Computer programs to overread mammograms and Pap smears have been available commercially for a number of years. These programs point out for the radiologists and cytopathologists suspicious areas that might have been overlooked. After some early studies with positive results that led to approval by the US Food and Drug Administration (FDA), these programs have been commercially available. Now that they have been in use for awhile, however, recently published, large-scale, randomized trials of both programs have raised doubts about their performance in practice.[193–195] A recently completed randomized trial of Pap smear results showed a very slight advantage of the computer programs over unaided cytopathologists,[194] but earlier reports of the trial before completion did not show any differences.[193] The authors suggest that it may take time for optimal quality to be achieved with a new technique.

In the area of computer-assisted mammography interpretation, a randomized trial showed no difference in cancer detection but an increase in false-positives with the use of the software compared with unaided interpretation by radiologists.[195] It is certainly possible that technical improvements have made later systems better than earlier ones, and, as suggested by Nieminen and colleagues[194] about the Pap smear program, and Hall[196] about the mammography programs, it may take time, perhaps years, for the users to learn how to properly interpret and work with the software. These results highlight that realizing the potential advantages of second opinions (human or automated) may be a challenge.

***Autopsy.*** Sir William Osler championed the belief that medicine should be learned from patients, at the bedside and in the autopsy suite. This approach was espoused by Richard Cabot and many others, a tradition that continues today in the "Clinical Pathological Correlation" (CPC) exercises published weekly in *The New England Journal of Medicine*. Autopsies and CPCs teach more than just the specific medical content; they also illustrate the uncertainty that is inherent in the practice of medicine and effectively convey the concepts of fallibility and diagnostic error.

Unfortunately, as discussed above, autopsies in the United States have largely disappeared. Federal tracking of autopsy rates was suspended a decade ago, at which point the autopsy rate had already fallen to <7%. Most trainees in medicine today will never see an autopsy. Patient safety advocates have pleaded to resurrect the autopsy as an effective tool to improve calibration and reduce overconfidence, but so far to no avail.[144,197]

If autopsies are not generally available, has any other process emerged to provide a comparable feedback experience? An innovative candidate is the "Morbidity and Mortality (M & M) Rounds on the Web" program sponsored by the Agency for Healthcare Research and Quality (AHRQ).[198] This site features a quarterly set of 4 cases, each involving a medical error. Each case includes a comprehensive, well-referenced discussion by a safety expert. These cases are attractive, capsulized gems that, like an autopsy, have the potential to educate clinicians regarding medical error, including diagnostic error. The unknown factor regarding this endeavor is whether these lessons will provide the same impact as an autopsy, which teaches by the principle of learning from one's own mistakes.[78] Local "morbidity and mortality" rounds have the same potential to alert providers to the possibility of error, and the impact of these exercises increases if the patient sustains harm.[199]

A final option to provide feedback in the absence of a formal autopsy involves detailed postmortem magnetic resonance imaging scanning. This option obviates many of the traditional objections to an autopsy, and has the potential to reveal many important diagnostic discrepancies.[200]

***Feedback in Other Field Settings (The Questec Experiment).*** A fascinating experiment is underway that could substantially clarify the power of feedback to improve calibration and performance. This is the Questec experiment sponsored by Major League Baseball to improve the consistency of umpires in calling balls and strikes. Questec is a company that installs cameras in selected stadiums that track the ball path across home plate. At the end of the game, the umpire is provided a recording that replays every pitch, and gives him the opportunity to compare the called balls and strikes with the true ball path.[201] Umpires have vigorously objected to this project, including a planned civil lawsuit to stop the experiment. The results from this study have yet to be released, but they will certainly shed light on the question of whether a skeptical cohort of professionals can improve their performance through directed feedback.

**Follow-up.** A systems approach recommended by Redelmeier[76] and Gandhi et al[77] is to promote the use of follow-up. Schiff[31,75] also has long advocated the importance of follow-up and tracking to improve diagnoses. Planned follow-up after the initial diagnosis allows time for other thoughts to emerge, and time for the clinician to apply more conscious problem-solving strategies (such as decision-support tools) to the problem. A very appealing aspect of planned follow-up is that a patient's problems will evolve over the intervening period, and these changes will either support the original diagnostic possibilities, or point toward alternatives. If the follow-up were done soon enough, this approach might also mitigate the potential harm of diagnostic error, even without solving the problem of how to prevent cognitive error in the first place.

## ANALYSIS OF STRATEGIES TO REDUCE OVERCONFIDENCE

The strategies suggested above, even if they are successful in addressing the problem of overconfidence or miscalibration, have limitations that must be acknowledged. One involves the trade-offs of time, cost, and accuracy. We can be more certain, but at a price.[202] A second problem is unanticipated negative effects of the intervention.

### Tradeoffs in Time, Cost, and Accuracy

As clinicians improve their diagnostic competency from beginning level skills to expert status, reliability and accuracy improve with decreased cost and effort. However, using the strategies discussed earlier to move nonexperts into the realm of experts will involve some expense. In any given case, we can improve diagnostic accuracy but with increased cost, time, or effort.

Several of the interventions entail direct costs. For instance, expenditures may be in the form of payment for consultation or purchasing diagnostic decision-support systems. Less tangible costs relate to clinician time. Attending training programs involves time, effort, and money. Even strategies that do not have direct expenses may still be costly in terms of physician time. Most medical decision making takes place in the "adaptive subconscious." The application of expert knowledge, pattern and script recognition, and heuristic synthesis takes place essentially instantaneously for the vast majority of medical problems. The process is effortless. If we now ask physicians to reflect on how they arrived at a diagnosis, the extra time and effort required may be just enough to discourage this undertaking.

Applying conscious review of subconscious processing hopefully uncovers at least some of the hidden biases that affect subconscious decisions. The hope is that these events outnumber the new errors that may evolve as we second-guess ourselves. However, it is not clear that conscious articulation of the reasoning process is an accurate picture of what really occurs in expert decision making. As discussed above, even reviewing the suggestions from a decision-support system (which would facilitate reflection) is perceived as taking too long, even though the information is viewed as useful.[173] Although these arguments may not be persuasive to the individual patient,[2] it is clear that the time involved is a barrier to physician use of decision aids. Thus, in deciding to use methods to increase reflection, decisions must be made as to: (1) whether the marginal improvements in accuracy are worth the time and effort and, given the extra time involved, (2) how to ensure that clinicians will routinely make the effort.

## Unintended Consequences

Innovations made in the name of improving safety sometimes create new opportunities to fail, or have unintended consequences that decrease the expected benefit. In this framework, we should carefully examine the possibility that some of the interventions being considered might actually increase the risk of diagnostic error.

As an example, consider the interventions we have grouped under the general heading of "reflective practice." Most of the education and feedback efforts, and even the consultation strategies, are aimed at increasing such reflection. Imagine a physician who has just interviewed and examined an elderly patient with crampy abdominal pain, and who has concluded that the most likely explanation is constipation. What is the downside of consciously reconsidering this diagnosis before taking action?

**It Takes More Time.** The extra time the reflective process takes not only affects the physician but may have an impact on the patient as well. The extra time devoted to this activity may actually delay the diagnosis for one patient and may be time subtracted from another.

**It Can Lead to Extra Testing.** As other possibilities are envisioned, additional tests and imaging may be ordered. Our patient with simple constipation now requires an abdominal CT scan. This greatly increases the chances of discovering incidental findings and the risk of inducing *cascade effects*, where one thing leads to another, all of them extraneous to the original problem.[203] Not only might these pose additional risks to the patient, such testing is also likely to increase costs.[173] The risk of changing a "right" diagnosis to a "wrong" one will necessarily increase as the number of options enlarges; research has found that this sometimes occurs in experimental settings.[99,168]

**It May Change the Patient-Physician Dynamic.** Like physicians, most patients much prefer certainty over ambiguity. Patients want to believe that their healthcare providers know exactly what their disorder is, and what to do about it. An approach that lays out all the uncertainties involved and the probabilistic nature of medical decisions is unlikely to be warmly received by patients unless they are highly sophisticated. A patient who is reassured that he or she most likely has constipation will probably sleep a lot better than the one who is told that the abdominal CT scan is needed to rule out more serious concerns.

**The Risk of Diagnostic Error May Actually Increase.** The quality of automatic decision making may be degraded if subjected to conscious inspection. As pointed out in *Blink*,[127] we can all easily envision Marilyn Monroe, but would be completely stymied in attempting to describe her well enough for a stranger to recognize her from a set of pictures. There is, in fact, evidence that complex decisions are solved best without conscious attention.[204] A complementary observation is that the quality of conscious decision making degrades as the number of options to be considered increases.[205]

**Increased Reliance on Consultative Systems May Result in "Deskilling."** Although currently the diagnostic decision-support systems claim that they are only providing suggestions, not "the definitive diagnosis,"[206]there is a tendency on the part of users to believe the computer. Tsai and colleagues[207] found that residents reading electrocardiograms improved their interpretations when the computer interpretation was correct, but were worse when it was incorrect. A study by Galletta and associates[208] using the spell-checker in a word-processing program found similar results. There is a risk that, as the automated programs get more accurate, users will rely on them and lose the ability to tell when the systems are incorrect.

A summary of the strategies, their assumptions, which may not always be accurate, and the tradeoffs in implementing them is shown in Table 2.

## RECOMMENDATIONS FOR FUTURE RESEARCH

> *"Happy families are all alike; every unhappy family is unhappy in its own way."*
>
> —Leo Tolstoy, *Anna Karenina*[209]

We are left with the challenge of trying to consider solutions based on our current understanding of the research

**Table 2** Strategies to Reduce Diagnostic Errors

| Strategy | Purpose | Timing | Focus | Underlying Assumptions | Tradeoffs |
|---|---|---|---|---|---|
| **Education and training** | | | | | |
| Training in reflective practice and avoidance of biases | Provide metacognitive skills | Not tied to specific patient cases | Individual, prevention | Transfer from educational to practice setting will occur; clinician will recognize when thinking is incorrect | Not tied to action: expensive and time consuming except in defined educational settings |
| Increase expertise | Provide knowledge and experience | Not tied to specific patient cases | Individual, prevention | Transfer across cases will occur; errors are a result of lack of knowledge or experience | Expensive and time consuming except in defined educational settings |
| **Consultation** | | | | | |
| Computer-based general knowledge resources | Validate or correct initial diagnosis; suggest alternatives | At the point-of-care while considering diagnosis | Individual, prevention | Users will recognize the need for information and will use the feedback provided | Delay in action; most sources still need better indexing to improve speed of accessing information |
| Second opinions/consult with experts | Validate or correct initial diagnosis | Before treatment of specific patient | System, prevention/mitigation | Expert is correct and/or agreement would mean diagnosis is correct | Delay in action; expense, bottlenecks, may need 3rd opinion if there is disagreement; if not mandatory would be only used for cases where physician is puzzled |
| DDSS | Validate or correct initial diagnosis | Before definitive diagnosis of specific patient | System, prevention | DDSS suggestions would include correct diagnosis; physician will recognize correct diagnosis when DDSS suggests it | Delay in action, cost of system; if not mandatory for all cases would be only used for cases where physician is puzzled |
| **Feedback** | | | | | |
| Increase number of autopsies/M&M | Prevent future errors | After an adverse event or death has occurred | System, prevention in future | Clinician will learn from errors and will not make them again; feedback will improve calibration | Cannot change action, too late for specific patient, expensive |
| Audit and feedback | Prevent future errors | At regular intervals covering multiple patients seen over a given period | System, prevention in future | Clinician will learn from errors and will not make them again; feedback will improve calibration | Cannot change action, too late for specific patient, expensive |
| Rapid follow-up | Prevent future errors and mitigate harm from errors for specific patient | At specified intervals unique to specific patients shortly after diagnosis or treatment | System, mitigation | Error may not be preventable, but harm in selected cases may be mitigated; feedback will improve calibration | Expense, change in workflow, MD time in considering problem areas |

DDSS = diagnostic decision-support system; MD = medical doctor; M&M = morbidity and mortality.

on overconfidence and the strategies to overcome it. Studies show that experts seem to know what to do in a given situation and what they know works well most of the time. What this means is that diagnoses are correct most of the time. However, as advocated in the Institute of Medicine (IOM) reports, the engineering principle of "design for the usual, but plan for the unusual" should apply to this situation.[210] As Gladwell[211] discussed in an article in *The New Yorker* on homelessness, however, the solutions to address the "unusual" (or the "unhappy families" referenced in the epigraph above) may be very different from those that work for the vast majority of cases. So while we are not advocating complacency in the face of error, we are assuming that some errors will escape our prevention. For these situations, we must have contingency plans in place for reducing the harm ensuing from them.

If we look at the aspects of overconfidence discussed in this review, the cognitive and systemic factors appear to be more easily addressed than the attitudinal issues and those related to complacency. However, the latter two may be affected by addressing the former ones. If physicians were better calibrated, i.e., knew accurately when they were correct or incorrect, arrogance and complacency would not be a problem.

Our review demonstrates that while all of the methods to reduce diagnostic error can potentially reduce misdiagnosis, none of the educational approaches are systematically used outside the initial educational setting and when automated devices operate in the background they are not used uniformly. Our review also shows that on some level, physicians' overconfidence in their own diagnoses and complacency in the face of diagnostic error can account for the lack of use. That is, given information and incentives to examine and modify one's initial diagnoses, physicians choose not to undertake the effort. Given that physicians in general are reasonable individuals, the only feasible explanation is that they believe that their initial diagnoses are correct (even when they are not) and there is no reason for change. We return to the problem that prompted this literature review, but with a more focused research agenda to address the areas listed below.

### Overconfidence

Because most studies actually addressed overconfidence indirectly and usually in laboratory as opposed to real-life settings, we still do not know the prevalence of overconfidence in practice, whether it is the same across specialties, and what its direct role is in misdiagnosis.

### Preventability of Diagnostic Error

One of the glaring issues that is unresolved in the research to date is the extent to which diagnostic errors are preventable. The answer to this question will influence error-reduction strategies.

### Mitigating Harm

More research and evaluation of strategies that focus on mitigating the harm from the errors is needed. The research approach should include what Nolan has called "making the error visible."[164] Because these errors are likely the ones that have traditionally been unrecognized, focusing research on them can provide better data on how extensively they occur in routine practice. Most strategies for addressing diagnostic errors have focused on prevention; it is in the area of mitigation where the strategies are sorely lacking.

### Debiasing

Is instruction on cognitive error and cognitive forcing strategies effective at improving diagnosis? What is the best stage of medical education to introduce this training? Does it transfer from the training to the practice setting?

### Feedback

How much feedback do physicians get and how much do they need? What mechanisms can be constructed to get them more feedback on their own cases? What are the most effective ways to learn from the mistakes of others?

### Follow-up

How can planned follow-up of patient outcomes be encouraged and what approaches can be used for rapid follow-up to provide more timely feedback on diagnoses?

### Minimizing the Downside

Does conscious attention decrease the chances of diagnostic error or increase it? Can we think of ways to minimize the possibility that conscious attention to diagnosis may actually make things worse?

## CONCLUSIONS

Diagnostic error exists at an appreciable rate, ranging from $<5\%$ in the perceptual specialties up to 15% in most other areas of medicine. In this review, we have examined the possibility that overconfidence contributes to diagnostic error. Our review of the literature leads us to 2 main conclusions.

### Physicians Overestimate the Accuracy of Their Diagnoses

Overconfidence exists and is probably a trait of human nature—we all tend to overestimate our skills and abilities. Physicians' overconfidence in their decision making may simply reflect this tendency. Physicians come to trust the fast and frugal decision strategies they typically use. These strategies succeed so reliably that physicians can become complacent; the failure rate is minimal and errors may not come to their attention for a variety of reasons. Physicians acknowledge that diagnostic error exists, but seem to believe that the likelihood of error is less than it really is. They believe that they personally are unlikely to make a mistake. Indirect evidence of overconfidence emerges from the routine disregard that physicians show for tools that might be helpful. They rarely seek out feedback, such as autopsies, that would clarify their tendency to err, and they tend not to participate in other exercises that would provide independent information on their diagnostic accuracy. They disregard guidelines for diagnosis and treatment. They tend to ignore decision-support tools, even when these are readily accessible and known to be valuable when used.

### Overconfidence Contributes to Diagnostic Error

Physicians in general have well-developed metacognitive skills, and when they are uncertain about a case they typically devote extra time and attention to the problem and often request consultation from specialty experts. We believe many or most cognitive errors in diagnosis arise from the cases where they *are* certain. These are the cases where the problem appears to be routine and resembles similar cases that the clinician has seen in the past. In these situations, the metacognitive angst that exists in more challenging cases may not arise. Physicians may simply stop thinking about the case, predisposing them to all of the pitfalls that result from our cognitive “dispositions to respond.” They fail to consider other contexts or other diagnostic possibilities, and they fail to recognize the many inherent shortcomings that derive from heuristic thinking.

In summary, improving patient safety will ultimately require strategies that take into account the data from this review—why diagnostic errors occur, how they can be prevented, and how the harm that results can be reduced.

## ACKNOWLEDGMENTS

We are grateful to Paul Mongerson for encouragement and financial support of this research. The authors also appreciate the insightful comments of Arthur S. Elstein, PhD, on an earlier draft of this manuscript. We also appreciate the assistance of Muzna Mirza, MBBS, MSHI, Grace Garey, and Mary Lou Glazer in compiling the bibliography.

## AUTHOR DISCLOSURES

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

**Eta S. Berner, EdD**, has no financial arrangement or affiliation with a corporate organization or manufacturer of a product discussed in this article.

**Mark L. Graber, MD**, has no financial arrangement or affiliation with a corporate organization or manufacturer of a product discussed in this article.

## References

1. Lowry F. Failure to perform autopsies means some MDs “walking in a fog of misplaced optimism.” *CMAJ.* 1995;153:811–814.
2. Mongerson P. A patient’s perspective of medical informatics. *J Am Med Inform Assoc.* 1995;2:79–84.
3. Blendon RJ, DesRoches CM, Brodie M, et al. Views of practicing physicians and the public on medical errors. *N Engl J Med.* 2002;347:1933–1940.
4. YouGov survey of medical misdiagnosis. Isabel Healthcare–Clinical Decision Support System, 2005. Available at: http://www.isabelhealthcare.com. Accessed April 3, 2006.
5. Burroughs TE, Waterman AD, Gallagher TH, et al. Patient concerns about medical errors in emergency departments. *Acad Emerg Med.* 2005;23:57–64.
6. Tierney WM. Adverse outpatient drug events—a problem and an opportunity. *N Engl J Med.* 2003;348:1587–1589.
7. Norman GR, Coblentz CL, Brooks LR, Babcook CJ. Expertise in visual diagnosis: a review of the literature. *Acad Med.* 1992;67(suppl):S78–S83.
8. Foucar E, Foucar MK. Medical error. In: Foucar MK, ed. *Bone Marrow Patholog,* 2nd ed. Chicago: ASCP Press, 2001:76–82.
9. Fitzgerald R. Error in radiology. *Clin Radiol.* 2001;56:938–946.
10. Kronz JD, Westra WH, Epstein JI. Mandatory second opinion surgical pathology at a large referral hospital. *Cancer.* 1999;86:2426–2435.
11. Berlin L, Hendrix RW. Perceptual errors and negligence. *Am J Radiol.* 1998;170:863–867.
12. Kripalani S, Williams MV, Rask K. Reducing errors in the interpretation of plain radiographs and computed tomography scans. In: Shojania KG, Duncan BW, McDonald KM, Wachter RM, eds. *Making Health Care Safer. A Critical Analysis of Patient Safety Practices*. Rockville, MD: Agency for Healthcare Research and Quality, 2001.
13. Neale G, Woloschynowych J, Vincent C. Exploring the causes of adverse events in NHS hospital practice. *J R Soc Med.* 2001;94:322–330.
14. O’Connor PM, Dowey KE, Bell PM, Irwin ST, Dearden CH. Unnecessary delays in accident and emergency departments: do medical and surgical senior house officers need to vet admissions? *Acad Emerg Med.* 1995;12:251–254.
15. Chellis M, Olson JE, Augustine J, Hamilton GC. Evaluation of missed diagnoses for patients admitted from the emergency department. *Acad Emerg Med.* 2001;8:125–130.
16. Elstein AS. Clinical reasoning in medicine. In: Higgs JJM, ed. *Clinical Reasoning in the Health Professions*. Oxford, England: Butterworth-Heinemann Ltd, 1995:49–59.
17. Kedar I, Ternullo JL, Weinrib CE, Kelleher KM, Brandling-Bennett H, Kvedar JC. Internet based consultations to transfer knowledge for patients requiring specialised care: retrospective case review. *BMJ.* 2003;326:696–699.
18. McGinnis KS, Lessin SR, Elder DE. Pathology review of cases presenting to a multidisciplinary pigmented lesion clinic. *Arch Dermatol.* 2002;138:617–621.
19. Zarbo RJ, Meier FA, Raab SS. Error detection in anatomic pathology. *Arch Pathol Lab Med.* 2005;129:1237–1245.
20. Tomaszewski JE, Bear HD, Connally JA, et al. Consensus conference on second opinions in diagnostic anatomic pathology: who, what, and when. *Am J Clin Pathol.* 2000;114:329–335.
21. Harris M, Hartley AL, Blair V, et al. Sarcomas in north west England. I. Histopathological peer review. *Br J Cancer.* 1991;64:315–320.
22. Kim J, Zelman RJ, Fox MA, et al. Pathology Panel for Lymphoma Clinical Studies: a comprehensive analysis of cases accumulated since its inception. *J Natl Cancer Inst.* 1982;68:43–67.
23. Goddard P, Leslie A, Jones A, Wakeley C, Kabala J. Error in radiology. *Br J Radiol.* 2001;74:949–951.
24. Berlin L. Defending the ”missed” radiographic diagnosis. *Am J Radiol.* 2001;176:317–322.

25. Espinosa JA, Nolan TW. Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study. *BMJ.* 2000; 320:737–740.
26. Arenson RL. The wet read. AHRQ [Agency for Heathcare Research and Quality] Web M&M, March 2006. Available at: http://webmm.ahrq.gov/printview.aspx?caseID=121. Accessed November 28, 2007.
27. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists: findings from a national sample. *Arch Intern Med.* 1996;156:209–213.
28. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. *Radiographics.* 2003; 23:881–895.
29. Goodson WH III, Moore DH II. Causes of physician delay in the diagnosis of breast cancer. *Arch Intern Med.* 2002;162:1343–1348.
30. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United Kingdom. *JAMA.* 2003;290:2129–2137.
31. Schiff GD, Kim S, Abrams R, et al. Diagnosing diagnosis 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 and Quality, February 2005. AHRQ Publication No. 050021. Available at: http://www.ahrq.gov/downloads/pub/advances/vol2/schiff.pdf./. Accessed December 3, 2007.
32. Shojania K, Burton E, McDonald K, et al. The autopsy as an outcome and performance measure: evidence report/technology assessment #58. Rockville, MD: Agency for Healthcare Research and Quality, October 2002. AHRQ Publication No. 03-E002.
33. Pidenda LA, Hathwar VS, Grand BJ. Clinical suspicion of fatal pulmonary embolism. *Chest.* 2001;120:791–795.
34. Lederle FA, Parenti CM, Chute EP. Ruptured abdominal aortic aneurysm: the internist as diagnostician. *Am J Med.* 1994;96:163–167.
35. von Kodolitsch Y, Schwartz AG, Nienaber CA. Clinical prediction of acute aortic dissection. *Arch Intern Med.* 2000;160:2977–2982.
36. Edlow JA. Diagnosis of subarachnoid hemorrhage. *Neurocrit Care.* 2005;2:99–109.
37. Burton EC, Troxclair DA, Newman WP III. Autopsy diagnoses of malignant neoplasms: how often are clinical diagnoses incorrect? *JAMA.* 1998;280:1245–1248.
38. Perlis RH. Misdiagnosis of bipolar disorder. *Am J Manag Care.* 2005;11(suppl):S271–S274.
39. Graff L, Russell J, Seashore J, et al. False-negative and false-positive errors in abdominal pain evaluation: failure to diagnose acute appendicitis and unneccessary surgery. *Acad Emerg Med.* 2000;7:1244–1255.
40. Raab SS, Grzybicki DM, Janosky JE, et al. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer.* 2005;104:2205–2213.
41. Buchweitz O, Wulfing P, Malik E. Interobserver variability in the diagnosis of minimal and mild endometriosis. *Eur J Obstet Gynecol Reprod Biol.* 2005;122:213–217.
42. Gorter S, van der Heijde DM, van der Linden S, et al. Psoriatic arthritis: performance of rheumatologists in daily practice. *Ann Rheum Dis.* 2002;61:219–224.
43. Bogun F, Anh D, Kalahasty G, et al. Misdiagnosis of atrial fibrillation and its clinical consequences. *Am J Med.* 2004;117:636–642.
44. Arnon SS, Schecter R, Maslanka SE, Jewell NP, Hatheway CL. Human botulism immune globulin for the treatment of infant botulism. *N Engl J Med.* 2006;354:462–472.
45. Edelman D. Outpatient diagnostic errors: unrecognized hyperglycemia. *Eff Clin Pract.* 2002;5:11–16.
46. Russell NJ, Pantin CF, Emerson PA, Crichton NJ. The role of chest radiography in patients presenting with anterior chest pain to the Accident & Emergency Department. *J R Soc Med.* 1988;81:626–628.
47. Dobbs D. Buried answers. *New York Times Magazine.* April 24, 2005:40–45.
48. Cabot RC. Diagnostic pitfalls identified during a study of three thousand autopsies. *JAMA.* 1912;59:2295–2298.
49. Cabot RC. A study of mistaken diagnosis: based on the analysis of 1000 autopsies and a comparison with the clinical findings. *JAMA.* 1910;55:1343–1350.
50. Aalten CM, Samsom MM, Jansen PA. Diagnostic errors: the need to have autopsies. *Neth J Med.* 2006;64:186–190.
51. Shojania KG. Autopsy revelation. AHRQ [Agency for Heathcare Research and Quality] Web M&M, March 2004. Available at: http://webmm.ahrq.gov/case.aspx?caseID=54&searchStr=shojania. Accessed November 28, 2007.
52. Tamblyn RM. Use of standardized patients in the assessment of medical practice. *CMAJ.* 1998;158:205–207.
53. Berner ES, Houston TK, Ray MN, et al. Improving ambulatory prescribing safety with a handheld decision support system: a randomized controlled trial. *J Am Med Inform Assoc*. 2006;13:171–179.
54. Christensen-Szalinski JJ, Bushyhead JB. Physician's use of probabalistic information in a real clinical setting. *J Exp Psychol Hum Percept Perform.* 1981;7:928–935.
55. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. *Med Care.* 2004;42:1066–1072.
56. Margo CE. A pilot study in ophthalmology of inter-rater reliability in classifying diagnostic errors: an underinvestigated area of medical error. *Qual Saf Health Care.* 2003;12:416–420.
57. Hoffman PJ, Slovic P, Rorer LG. An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. *Psychol Bull.* 1968;69:338–349.
58. Kohn L, Corrigan JM, Donaldson M. *To Err Is Human: Building a Safer Health System.* Washington, DC: National Academy Press, 1999.
59. Leape L, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II. *N Engl J Med.* 1991;324:377–384.
60. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care.* 2000;38:261–271.
61. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ.* 2004;170:1678–1686.
62. Wilson RM, Harrison BT, Gibberd RW, Hamilton JD. An analysis of the causes of adverse events from the Quality in Australian Health Care Study. *Med J Aust.* 1999;170:411–415.
63. Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals II: preventability and clinical context. *N Z Med J.* 2003;116:U624.
64. Bhasale A, Miller G, Reid S, Britt HC. Analyzing potential harm in Australian general practice: an incident-monitoring study. *Med J Aust.* 1998;169:73–76.
65. Makeham M, Dovey S, County M, Kidd MR. An international taxonomy for errors in general practice: a pilot study. *Med J Aust.* 2002;177:68–72.
66. Fischer G, Fetters MD, Munro AP, Goldman EB. Adverse events in primary care identified from a risk-management database. *J Fam Pract.* 1997;45:40–46.
67. Wu AW, Folkman S, McPhee SJ, Lo B. Do house officers learn from their mistakes? *JAMA.* 1991;265:2089–2094.
68. Weingart S, Ship A, Aronson M. Confidential clinical-reported surveillance of adverse events among medical inpatients. *J Gen Intern Med.* 2000;15:470–477.
69. Balsamo RR, Brown MD. Risk management. In: Sanbar SS, Gibofsky A, Firestone MH, LeBlang TR, eds. *Legal Medicin*e, 4th ed. St Louis, MO: Mosby, 1998:223–244.
70. Failure to diagnose. Midiagnosis of conditions and diseaes. Medical Malpractice Lawyers and Attorneys Online, 2006. Available at: http://www.medical-malpractice-attorneys-lawsuits.com/pages/failure-to-diagnose.html. Accessed November 28, 2007.

71. *General and Family Practice Claim Summary.* Physician Insurers Association of America, Rockville, MD, 2002.
72. Berlin L. Fear of cancer. *AJR Am J Roentgenol.* 2004;183:267–272.
73. Missed or failed diagnosis: what the UNITED claims history can tell us. United GP Registrar's Toolkit, 2005. Available at: http://www.unitedmp.com.au/0/0.13/0.13.4/Missed_diagnosis.pdf. Accessed November 28, 2007.
74. Studdert DM, Mello MM, Gawande AA, et al. Claims, errors, and compensation payments in medical malpractice litigation. *N Engl J Med*. 2006;354:2024–2033.
75. Schiff GD. Commentary: diagnosis tracking and health reform. *Am J Med Qual.* 1994;9:149–152.
76. Redelmeier DA. Improving patient care: the cognitive psychology of missed diagnoses. *Ann Intern Med.* 2005;142:115–120.
77. 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.
78. Kirch W, Schafii C. Misdiagnosis at a university hospital in 4 medical eras. *Medicine (Baltimore).* 1996;75:29–40.
79. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three different eras. *N Engl J Med.* 1983;308:1000–1005.
80. 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.
81. Sonderegger-Iseli K, Burger S, Muntwyler J, Salomon F. Diagnostic errors in three medical eras: a necropsy study. *Lancet.* 2000;355: 2027–2031.
82. Berner ES, Miller RA, Graber ML. Missed and delayed diagnoses in the ambulatory setting. *Ann Intern Med.* 2007;146:470–471.
83. Gawande A. Final cut. Medical arrogance and the decline of the autopsy. *The New Yorker.* March 19, 2001:94–99.
84. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999;77:1121–1134.
85. LaFee S. Well news: all the news that's fit. *The San Diego Union-Tribune.* March 7, 2006. Available at: http://www.quotegarden.com/medical.html. Accessed February 6, 2008.
86. Graber ML. Diagnostic error in medicine: a case of neglect. *Jt Comm J Qual Patient Saf.* 2005;31:112–119.
87. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? *Ann Intern Med.* 1985;103:596–599.
88. Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. *Med Decis Making.* 1995;15:113–119.
89. Osheroff JA, Bankowitz RA. Physicians' use of computer software in answering clinical questions. *Bull Med Libr Assoc.* 1993;81:11–19.
90. Rosenbloom ST, Geissbuhler AJ, Dupont WD, et al. Effect of CPOE user interface design on user-initiated access to educational and patient information during clinical care. *J Am Med Inform Assoc.* 2005;12:458–473.
91. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med.* 2003;348: 2635–2645.
92. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA.* 1999;282:1458–1465.
93. Eccles MP, Grimshaw JM. Selecting, presenting and delivering clinical guidelines: are there any "magic bullets"? *Med J Aust.* 2004; 180(suppl):S52–S54.
94. Pearson TA, Laurora I, Chu H, Kafonek S. The lipid treatment assessment project (L-TAP): a multicenter survey to evaluate the percentages of dyslipidemic patients receiving lipid-lowering therapy and achieving low-density lipoprotein cholesterol goals. *Arch Intern Med.* 2000;160:459–467.
95. Eccles M, McColl E, Steen N, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial [primary care]. *BMJ.* 2002;325:941.
96. Smith WR. Evidence for the effectiveness of techniques to change physician behavior. *Chest.* 2000;118:8S–17S.
97. Militello L, Patterson ES, Tripp-Reimer T, et al. Clinical reminders: why don't people use them? In: *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, New Orleans LA, 2004:1651–1655.
98. Patterson ES, Doebbeling BN, Fung CH, Militello L, Anders S, Asch SM. Identifying barriers to the effective use of clinical reminders: bootstrapping multiple methods. *J Biomed Inform.* 2005;38:189–199.
99. Berner ES, Maisiak RS, Heudebert GR, Young KR Jr. Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. *AMIA Annu Symp Proc.* 2003; 2003:76–80.
100. Steinman MA, Fischer MA, Shlipak MG, et al. Clinician awareness of adherence to hypertension guidelines. *Am J Med.* 2004;117:747–754.
101. Tierney WM, Overhage JM, Murray MD, et al. Can computer-generated evidence-based care suggestions enhance evidence-based management of asthma and chronic obstructive pulmonary disease? A randomized, controlled trial. *Health Serv Res.* 2005;40:477–497.
102. Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine. *Health Aff (Millwood).* 2005;24:18–28.
103. Tanenbaum SJ. Evidence and expertise: the challenge of the outcomes movement to medical professionalism. *Acad Med.* 1999;74: 757–763.
104. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc.* 2006;13:138–147.
105. Katz J. Why doctors don't disclose uncertainty. *Hastings Cent Rep.* 1984;14:35–44.
106. Graber ML, Franklin N, Gordon RR. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165:1493–1499.
107. Croskerry P. Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Acad Emerg Med.* 2002;9:1184–1204.
108. Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? *J Gen Intern Med.* 2005;20:334–339.
109. Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artif Intell Med.* 2005;33:25–30.
110. Baumann AO, Deber RB, Thompson GG. Overconfidence among physicians and nurses: the 'micro-certainty, macro-uncertainty' phenomenon. *Soc Sci Med.* 1991;32:167–174.
111. Podbregar M, Voga G, Krivec B, Skale R, Pareznik R, Gabrscek L. Should we confirm our clinical diagnostic certainty by autopsies? *Intensive Care Med.* 2001;27:1750–1755.
112. Landefeld CS, Chren MM, Myers A, Geller R, Robbins S, Goldman L. Diagnostic yield of the autopsy in a university hospital and a community hospital. *N Engl J Med.* 1988;318:1249–1254.
113. Potchen EJ. Measuring observer performance in chest radiology: some experiences. *J Am Coll Radiol.* 2006;3:423–432.
114. Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med.* 2007;49:196–205.
115. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med.* 2003;78:775–780.
116. Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. *J Eval Clin Pract.* 2001;7:97–107.
117. McSherry D. Avoiding premature closure in sequential diagnosis. *Artif Intell Med.* 1997;10:269–283.
118. Dubeau CE, Voytovich AE, Rippey RM. Premature conclusions in the diagnosis of iron-deficiency anemia: cause and effect. *Med Decis Making.* 1986;6:169–173.
119. Voytovich AE, Rippey RM, Suffredini A. Premature conclusions in diagnostic reasoning. *J Med Educn.* 1985;60:302–307.

120. Simon HA. *The Sciences of the Artificial,* 3rd ed. Cambridge, MA: MIT Press, 1996.
121. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving. An Analysis of Clinical Reasoning.* Cambridge, MA: Harvard University Press, 1978.
122. Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning of randomly selected physicians in general medical practice. *Clin Invest Med.* 1982;5:49–55.
123. Barrows HS, Feltovich PJ. The clinical reasoning process. *Med Educ.* 1987;21:86–91.
124. Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Med Educ.* 1981;15:315–322.
125. Norman GR. The epistemology of clinical reasoning: perspectives from philosophy, psychology, and neuroscience. *Acad Med.* 2000; 75(suppl):S127–S135.
126. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. *Acad Med.* 1990;65: 611–621.
127. Gladwell M. *Blink: The Power of Thinking Without Thinking*. Boston: Little Brown and Company, 2005.
128. Klein G. *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press, 1998.
129. Rosch E, Mervis CB. Family resemblances: studies in the internal structure of categories. *Cognit Psychol.* 1975;7:573–605.
130. Eva KW, Norman GR. Heuristics and biases—a biased perspective on clinical reasoning. *Med Educ.* 2005;39:870–872.
131. Sieck WR, Arkes HR. The recalcitrance of overconfidence and its contribution to decision aid neglect. *J Behav Decis Making.* 2005; 18:29–53.
132. Kruger J, Dunning D. Unskilled and unaware—but why? A reply to Krueger and Mueller (2002). *J Pers Soc Psychol.* 2002;82:189–192.
133. Krueger J, Mueller RA. Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *J Pers Soc Psychol.* 2002;82:180–188.
134. Mele AR. Real self-deception. *Behav Brain Sci.* 1997;20:91–102.
135. Reason JT, Manstead ASR, Stradling SG. Errors and violation on the roads: a real distinction? *Ergonomics.* 1990;33:1315–1332.
136. Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. *Psychol Rev.* 1996;103:650–669.
137. Hamm RM. Clinical intuition and clinical analysis: expertise and the cognitive continuum. In: Elstein A, Dowie J, eds. *Professional Judgment: A Reader in Clinical Decision Making*. Cambridge, UK: Cambridge University Press, 1988:78–105.
138. Gigerenzer G. *Adaptive Thinking.* New York: Oxford University Press, 2000.
139. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. *N Engl J Med.* 1994;330:1792–1796.
140. Ely JW, Levinson W, Elder NC, Mainous AG III, Vinson DC. Perceived causes of family physicians' errors. *J Fam Pract.* 1995; 40:337–344.
141. Studdert DM, Mello MM, Sage WM, et al. Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *JAMA.* 2005;293:2609–2617.
142. Anderson RE. Billions for defense: the pervasive nature of defensive medicine. *Arch Intern Med.* 1999;159:2399–2402.
143. Trivers R. The elements of a scientific theory of self-deception. *Ann N Y Acad Sci.* 2000;907:114–131.
144. Lundberg GD. Low-tech autopsies in the era of high-tech medicine: continued value for quality assurance and patient safety. *JAMA.* 1998;280:1273–1274.
145. Darwin C. *The Descent of Man.* Project Gutenberg, August 1, 2000. Available at: http://www.gutenberg.org/etext/2300. Accessed November 28, 2007.
146. Leonhardt D. Why doctors so often get it wrong. *The New York Times.* February 22, 2006 [published correction appears in *The New York Times*, February 28, 2006]. Available at: http://www.nytimes.com/2006/02/22/business/22leonhardt.html?ex=1298264400&en=c2d9f1d654850c17&ei=5088&partner=rssnyt&emc=rss. Accessed November 28, 2007.
147. Hodges B, Regehr G, Martin D. Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. *Acad Med.* 2001;76(suppl):S87–S89.
148. Davis DA, Mazmanian PE, Fordis M, Van HR, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA.* 2006;296:1094–1102.
149. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *JAMA.* 1999;282:867–874.
150. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med.* 2006;355:2217–2225.
151. Norman G. Building on experience—the development of clinical reasoning. *N Engl J Med.* 2006;355:2251–2252.
152. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med.* 1999;74(suppl):S128–S143.
153. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ.* 2005;39:418–427.
154. Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. *Qual Saf Health Care.* 2006;15: 159–164.
155. Croskerry P. When diagnoses fail: new insights, old thinking. *Canadian Journal of CME.* 2003;Nov:79–87.
156. Hall KH. Reviewing intuitive decision making and uncertainty: the implications for medical education. *Med Educ.* 2002;36:216–224.
157. Croskerry P. Cognitive forcing strategies in clinical decision making. *Ann Emerg Med.* 2003;41:110–120.
158. Mitchell DJ, Russo JE, Pennington N. Back to the future: temporal perspective in the explanation of events. *J Behav Decis Making.* 1989;2:25–38.
159. Schon DA. *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass, 1987.
160. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ.* 2004;38:1302–1308.
161. Soares SMS. Reflective practice in medicine (PhD thesis). Erasmus Universiteit, Rotterdam, Rotterdam, the Netherlands, 2006. 30552B 6000.
162. Mamede S, Schmidt HG, Rikers R. Diagnostic errors and reflective practice in medicine. *J Eval Clin Pract.* 2007;13:138–145.
163. Reason J. Human error: models and management. *BMJ.* 2000;320: 768–770.
164. Nolan TW. System changes to improve patient safety. *BMJ.* 2000; 320:771–773.
165. Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *J Am Med Inform Assoc.* 1994;1:8–27.
166. Kassirer JP. A report card on computer-assisted diagnosis—the grade: C. *N Engl J Med*. 1994;330:1824–1825.
167. Berner ES, Maisiak RS. Influence of case and physician characteristics on perceptions of decision support systems. *J Am Med Inform Assoc.* 1999;6:428–434.
168. Friedman CP, Elstein AS, Wolf FM, et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. *JAMA.* 1999;282:1851–1856.
169. Lincoln MJ, Turner CW, Haug PJ, et al. Iliad's role in the generalization of learning across a medical domain. *Proc Annu Symp Comput Appl Med Care.* 1992; 174–178.
170. Lincoln MJ, Turner CW, Haug PJ, et al. Iliad training enhances medical students' diagnostic skills. *J Med Syst.* 1991;15:93–110.
171. Turner CW, Lincoln MJ, Haug P, et al. Iliad training effects: a cognitive model and empirical findings. *Proc Annu Symp Comput Appl Med Care.* 1991:68–72.

172. Arene I, Ahmed W, Fox M, Barr CE, Fisher K. Evaluation of quick medical reference (QMR) as a teaching tool. *MD Comput.* 1998;15: 323–326.
173. Apkon M, Mattera JA, Lin Z, et al. A randomized outpatient trial of a decision-support information technology tool. *Arch Intern Med.* 2005;165:2388–2394.
174. Ramnarayan P, Roberts GC, Coren M, et al. Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: a quasi-experimental study. *BMC Med Inform Decis Mak.* 2006;6:22.
175. Maffei FA, Nazarian EB, Ramnarayan P, Thomas NJ, Rubenstein JS. Use of a web-based tool to enhance medical student learning in the pediatric ICU and inpatient wards. *Pediatr Crit Care Med.* 2005;6: 109.
176. Ramnarayan P, Tomlinson A, Kularni G, Rao A, Britto J. A novel diagnostic aid (ISABEL): development and preliminary evaluation of clinical performance. *Medinfo.* 2004;11:1091–1095.
177. Ramnarayan P, Kapoor RR, Coren J, et al. Measuring the impact of diagnostic decision support on the quality of clinical decision making: development of a reliable and valid composite score. *J Am Med Inform Assoc.* 2003;10:563–572.
178. Ramnarayan P, Tomlinson A, Rao A, Coren M, Winrow A, Britto J. ISABEL: a web-based differential diagnosis aid for paediatrics: results from an initial performance evaluation. *Arch Dis Child*. 2003; 88:408–413.
179. Miller RA. Evaluating evaluations of medical diagnostic systems. *J Am Med Inform Assoc.* 1996;3:429–431.
180. Berner ES. Diagnostic decision support systems: how to determine the gold standard? *J Am Med Inform Assoc.* 2003;10:608–610.
181. Bauer BA, Lee M, Bergstrom L, et al. Internal medicine resident satisfaction with a diagnostic decision support system (DXplain) introduced on a teaching hospital service. *Proc AMIA Symp.* 2002; 31–35.
182. Teich JM, Merchia PR, Schmiz JL, Kuperman GJ, Spurr CD, Bates DW. Effects of computerized physician order entry on prescribing practices. *Arch Intern Med.* 2000;160:2741–2747.
183. Croskerry P. The feedback sanction. *Acad Emerg Med.* 2000;7:1232–1238.
184. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care.* 2006;15:433–436.
185. Papa FJ, Aldrich D, Schumacker RE. The effects of immediate online feedback upon diagnostic performance. *Acad Med.* 1999;74(suppl): S16–S18.
186. Stone ER, Opel RB. Training to improve calibration and discrimination: the effects of performance and environment feedback. *Organ Behav Hum Decis Process.* 2000;83:282–309.
187. Duffy FD, Holmboe ES. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. *JAMA.* 2006;296:1137–1139.
188. Borgstede JP, Zinninger MD. Radiology and patient safety. *Acad Radiol.* 2004;11:322–332.
189. Frable WJ. "Litigation cells" in the Papanicolaou smear: extramural review of smears by "experts." *Arch Pathol Lab Med.* 1997;121:292–295.
190. Wilbur DC. False negatives in focused rescreening of Papanicolaou smears: how frequently are "abnormal" cells detected in retrospective review of smears preceding cancer or high grade intraepithelial neoplasia? *Arch Pathol Lab Med.* 1997;121:273–276.
191. College of American Pathologists. Available at: http://www.cap.org/apps/cap.portal.
192. Raab SS. Improving patient safety by examining pathology errors. *Clin Lab Med.* 2004;24:863.
193. Nieminen P, Kotaniemi L, Hakama M, et al. A randomised public-health trial on automation-assisted screening for cervical cancer in Finland: performance with 470,000 invitations. *Int J Cancer.* 2005; 115:307–311.
194. Nieminen P, Kotaniemi-Talonen L, Hakama M, et al. Randomized evaluation trial on automation-assisted screening for cervical cancer: results after 777,000 invitations. *J Med Screen.* 2007;14:23–28.
195. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med.* 2007;356:1399–1409.
196. Hall FM. Breast imaging and computer-aided detection. *N Engl J Med.* 2007;356:1464–1466.
197. Hill RB, Anderson RE. *Autopsy: Medical Practice and Public Policy*. Boston: Butterworth-Heinemann, 1988.
198. Cases and commentaries. AHRQ Web M&M, October 2007. Available at: http://www.webmm.ahrq.gov/index.aspx. Accessed December 12, 2007.
199. Fischer MA, Mazor KM, Baril J, Alper E, DeMarco D, Pugnaire M. Learning from mistakes: factors that influence how students and residents learn from medical errors. *J Gen Intern Med.* 2006;21:419–423.
200. Patriquin L, Kassarjian A, Barish M, et al. Postmortem whole-body magnetic resonance imaging as an adjunct to autopsy: preliminary clinical experience. *J Magn Reson Imaging*. 2001;13:277–287.
201. Umpire Information System (UIS). Available at: http://www.questec.com/q2001/prod_uis.htm. Accessed April 10, 2008.
202. Graber ML, Franklin N, Gordon R. Reducing diagnostic error in medicine: what's the goal? *Acad Med.* 2002;77:981–992.
203. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health.* 2002;23:23–44.
204. Dijksterhuis A, Bos MW, Nordgren LF, van Baaren RB. On making the right choice: the deliberation-without-attention effect. *Science.* 2006;311:1005–1007.
205. Redelmeier DA, Shafir E. Medical decision making in situations that offer multiple alternatives. *JAMA.* 1995;273:302–305.
206. Miller RA, Masarie FE Jr. The demise of the "Greek Oracle" model for medical diagnostic systems. *Methods Inf Med.* 1990;29:1–2.
207. Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. *J Am Med Inform Assoc*. 2003;10:478–483.
208. Galletta DF, Durcikova A, Everard A, Jones BM. Does spell-checking software need a warning label? *Communications of the ACM.* 2005;48:82–86.
209. Tolstoy L. *Anna Karenina*. Project Gutenberg, July 1, 1998. Available at: http://www.gutenberg.org/etext/1399. Accessed December 11, 2007.
210. Committee on Quality of Health Care in America, Institute of Medicine Report. Washington, DC: The National Academy Press, 2001.
211. Gladwell M. Million-dollar Murray. *The New Yorker.* February 13, 2006:96–107.

The American Journal of Medicine (2008) Vol 121 (5A), S24–S29

THE AMERICAN JOURNAL of MEDICINE®

# Overconfidence in Clinical Decision Making

Within medicine, there are more than a dozen major disciplines and a variety of further subspecialties. They have evolved to deal with >10,000 specific illnesses, all of which must be diagnosed before patient treatment can begin. This commentary is confined to orthodox medicine; diagnosis using folk and pseudo-diagnostic methods occurs in complementary and alternative medicine (CAM) and is described elsewhere.[1,2]

The process of developing an accurate diagnosis involves decision making. The patient typically enters the system through 1 of 2 portals: either the family doctor's office/a walk-in clinic or the emergency department. In both arenas, the first presentation of the illness is at its most undifferentiated. Often, the condition is diagnosed and treated, and the process ends there. Alternately, the general domain where the diagnosis probably lies is identified and the patient is referred for further evaluation. Generally, uncertainty progressively decreases during the evaluative process. By the time the patient is in the hands of subspecialists, most of the uncertainty is removed. This is not to say that complete assurance ever prevails; in some areas (e.g., medicine, critical care, trauma, and surgery), considerable further diagnostic effort may be required due to the dynamic evolving nature of the patient's condition and further challenges arising during the course of management.

For the purposes of the present discussion, we can make a broad division of medicine into 2 categories: one that deals with most of the uncertainty about diagnosis (e.g., family medicine [FM] and emergency medicine [EM]) and the other wherein a significant part of the uncertainty is removed (e.g., the specialty disciplines). Internal medicine (IM) falls somewhere between the two in that diagnostic refinement is already underway but may be incomplete.

Benchmark studies in patient safety found that diagnostic failure was highest in FM, EM, and IM,[3-5] presumably reflecting the relatively high degree of diagnostic uncertainty. These settings, therefore, deserve the closest scrutiny. To examine this further, we need to look at the decision-making behaviors that underlie the diagnostic process, particularly the biases that may be involved. Overconfidence is one of the most significant of these biases. This paper expands on the article by Drs. Berner and Graber[6] in this supplement in regard to modes of diagnostic decision making and their relationship to the phenomenon of overconfidence.

## DUAL PROCESSING APPROACH TO DECISION MAKING

Effective problem solving, sound judgment, and well-calibrated clinical decision making are considered to be among the highest attributes of physicians. Surprisingly, however, this important area has been actively researched for only about 35 years. The main epistemological issues in clinical decision making have been reviewed.[7] Much current work in cognitive science suggests that the brain utilizes 2 subsystems for thinking, knowing, and information processing: System 1 and System 2.[8-12] Their characteristics are listed in **Table 1,** adapted from Hammond[9] and Stanovich.[13]

What is now known as System 1 corresponds to what Hammond[9] descibed as *intuitive*, referring to a decision mode that is dominated by heuristics such as mental shortcuts, maxims, and rules of thumb. The system is fast, associative, inductive, frugal, and often primed by an affective component. Importantly, our first reactions to any situation often have an affective valence.[14] Blushing, for example, is an unconscious response to specific situational stimuli. Though socially uncomfortable, it often is very revealing about deeper beliefs and conflicts. Generally, under conditions of uncertainty, we tend to trust these reflexive, associatively generated feelings.

Stanovich[13] adopted the term "the autonomous set of systems" (TASS), emphasizing the autonomous and reflexive nature of this style of responding to salient features of a situation (**Table 2**),[13] and providing further characterization of System 1 decision making. TASS is multifarious. It encompasses processes of emotional regulation and implicit learning. It also incorporates Fodorian modular theory,[15] which proposes that the brain has a variety of modules that have undergone Darwinian selection to deal with different contingencies of the immediate environment. TASS responses are, therefore, highly context bound. Importantly,

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

Requests for reprints should be addressed to Pat Croskerry, MD, PhD, Department of Emergency Medicine, Dalhousie University. 351 Bethune, VG Site, 1278 Tower Road, Halifax, Nova Scotia, Canada B3H 2Y9.

E-mail address: xkerry@ns.sympatico.ca.


doi:10.1016/j.amjmed.2008.02.001

**Table 1** Characteristics of System 1 and System 2 approaches in decision making

| Characteristic | System 1 (intuitive) | System 2 (analytic) |
|---|---|---|
| Cognitive style | Heuristic | Systematic |
| Operation | Associative | Rule based |
| Processing | Parallel | Serial |
| Cognitive awareness | Low | High |
| Conscious control | Low | High |
| Automaticity | High | Low |
| Rate | Fast | Slow |
| Reliability | Low | High |
| Errors | Normative distribution | Few but significant |
| Effort | Low | High |
| Predictive power | Low | High |
| Emotional valence | High | Low |
| Detail on judgment process | Low | High |
| Scientific rigor | Low | High |
| Context | High | Low |

Adapted from *Concise Encyclopedia of Information Processing in Systems and Organizations*,[9] and *The Robot's Rebellion: Finding Meaning in the Age of Darwin*.[13]

**Table 2** Properties of the autonomous set of systems (TASS)

- Processing takes place beyond conscious awareness
- Parallel processing: each "hard-wired" module can independently respond to the appropriate triggering stimulus, and more than 1 can respond at a time. Therefore, many different subprocesses can execute simultaneously
- An accepting system that does not consider opposites: tendency to focus only on what is true rather than what is false. Disposed to believe rather than take the skeptic position; therefore look to confirm rather than disconfirm (the analytic system, in contrast, is able to undo acceptance)
- Higher cognitive (intellectual) ability appears to be correlated with an ability to use System 2 to override TASS and produce responses that are instrumentally rational
- Typically driven by social, narrative and contextualizing styles, whereas the style of System 2 requires detachment, decoupling, and decontextualization

Adapted from *The Robot's Rebellion: Finding Meaning in the Age of Darwin*.[13]

repeated use of analytic (System 2) outputs can allow them to be relegated to the TASS level.[13]

Thus, the effortless pattern recognition that characterizes the clinical acumen of the expert physician is made possible by accretion of a vast experience (the repetitive use of a System 2 analytic approach) that eventually allows the process to devolve to an automatic level.[16,17] Indeed, it is the apparent effortlessness of the method that permits some disparaging discounting; physicians often refer to diagnosis based on System 1 thinking as "just pattern recognition." The process is viewed as simply a transition to an automatic way of thinking, analogous to that occurring in the variety of complex skills required for driving a car; eventually, after considerable practice, one arrives at the destination with little conscious recollection of the mechanisms for getting there.

The essential characteristic of this "nonanalytic" reasoning is that it is a process of matching the new situation to 1 of many exemplars in memory,[18] which are apparently retrievable rapidly and effortlessly. As a consequence, it may require no more mental effort for a clinician to recognize that the current patient is having a heart attack than it is for a child to recognize that a dog is a four-legged beast. This strategy of reasoning based on similarity to a prior learned example has been described extensively in the literature on exemplar models of concept formation.[19,20]

Overall, although generally adaptive and often useful for our purposes,[21,22] in some clinical situations, System 1 approaches may fail. When the signs and symptoms of a particular presentation do not fit into TASS, the response will not be triggered,[16] and recognition failure will result in System 2 being engaged instead. The other side of the coin is that occasionally people act against their better judgment and behave irrationally. Thus, it may be that under certain conditions, despite a rational judgment having been reached using System 2, the decision maker defaults to System 1. This is not an uncommon phenomenon in medicine; despite being aware of good evidence from painstakingly developed practice guidelines, clinicians may still overconfidently choose to follow their intuition.

In contrast, System 2 is *analytical*, i.e., deductive, slow, rational, rule based, and low in emotional investment. Unlike the hard-wired, parallel-processing capabilities of System 1, System 2 is a linear processor that follows explicit computational rules. It corresponds to the software of the brain, i.e., our learned, rational reasoning power. According to Stanovich,[13] this mode allows us "to sustain the powerful context-free mechanisms of logical thought, inference, abstraction, planning, decision making, and cognitive control."

Whereas it is natural to think that System 2 thinking—coldly logical and analytical—likely is superior to System 1, much depends on context. A series of studies[23,24] have shown that "pure" System 1 and System 2 thinking are error prone; a combination of the 2 is closer to optimal. A simple example suffices: the first time a student answers the question "what is 16 x 16?" System 2 thinking is used to compute slowly and methodically by long multiplication. If the question is posed again soon after, the student recognizes the solution and volunteers the answer quickly and accurately (assuming it was done correctly the first time) using System 1 thinking. Therefore, it is important for decision makers to be aware of which system they are using and its overall appropriateness to the situation.

Certain contexts do not allow System 1. We could not use this mode, for example, to put a man on the moon; only System 2 would have worked. In contrast, adopting an analytical System 2 approach in an emergent situation, where rapid decision making is called for, may be paradoxically irrational.[16] In this situation, the rapid cognitive style known popularly as "thin-slicing"[25] that characterizes System 1 might be more expedient and appropriate. Recent studies suggest that making unconscious snap decisions (*deliberation-without-attention effect*) can outperform more deliberate "rational" thinking in certain situations.[26,27]

Perhaps the mark of good decision makers is their ability to match Systems 1 and 2 to their respective optimal contexts and to consciously blend them into their overall decision making. Although TASS operates at an unconscious level, their output, once seen, can be consciously modulated by adding a System 2 approach. Engagement of System 2 may occur when it "catches" an error in System 1.[28]

## OVERCONFIDENCE

Overconfident judgment by clinicians is 1 example of many cognitive biases that may influence reasoning and medical decision making. This bias has been well demonstrated in the psychology literature, where it appears as a common, but not universal, finding.[29,30] Ethnic cross-cultural variations in overconfidence have been described.[31] Further, we appear to be consistently overconfident when we express extreme confidence.[29] Overconfidence also plays a role in self-assessment, where it is axiomatic that relatively incompetent individuals consistently overestimate their abilities.[32,33] In some circumstances, overconfidence would qualify as irrational behavior.

Why should overconfidence be a general feature of human behavior? First, this trait usually leads to definitive action, and cognitive evolutionists would argue that in our distant pasts definitive action, under certain conditions, would confer a selective advantage. For example, to have been certain of the threat of danger in a particular situation and to have acted accordingly increased the chances of that decision maker's genes surviving into the next generation. Equivocation might have spelled extinction. The "false alarm" cost (taking evasive action) was presumably minimal although some degree of signal-detection trade-off was necessary so that the false-positive rate was not too high and wasteful. Indeed, error management theory suggests that some cognitive biases have been selected due to such cost/benefit asymmetries for false-negative and false-positive errors.[34] Second, as has been noted, System 1 intuitive thinking may be associated with strong emotions such as excitement and enthusiasm. Such positive feelings, in turn, have been linked with an enhanced level of confidence in the decision maker's own judgment.[35] Third, from TASS perspective, it is easy to see how some individuals would be more accepting of, and overly confident in, apparent solutions and conclusions they have reached, rather than take a more skeptical stance and look for disconfirming evidence to challenge their assumptions. This is referred to as *confirmation bias*,[36] which is one of the most powerful of the cognitive biases. Not surprisingly, it takes far more mental effort to contemplate disconfirmation (the clinician can only be confident that something *isn't* disease A by considering all the other things it might be) than confirmation. Harkening back to the self-assessment literature, one can only assess how much one knows by accurately assessing how much one doesn't know, and as Frank Keil[37] says, "How can I know what I don't know when I don't know what I don't know?"

Importantly, overconfidence appears to be related to the amount and strength of supporting evidence people can find to support their viewpoints.[38] Thus, overconfidence itself may depend upon confirmation bias. People's judgments were better calibrated (there was less overconfidence) when they were obliged to take account of disconfirming evidence.[38] Hindsight bias appears to be another example of overconfidence; it could similarly be debiased by forcing a consideration of alternative diagnoses.[39] This *consider-the-opposite* strategy appears to be one of the more effective debiasing strategies. Overall, it appears to be the biased fashion in which evidence is generated during the development of a particular belief or hypothesis that leads to overconfidence.

Other issues regarding overconfidence may have their origins within the culture of medicine. Generally, it is considered a weakness and a sign of vulnerability for clinicians to appear unsure. Confidence is valued over uncertainty, and there is a prevailing censure against disclosing uncertainty to patients.[40] There are good reasons for this. Shamans, the progenitors of modern clinicians, would have suffered short careers had they equivocated about their cures.[41] In the present day, the charisma of physicians and the confidence they have in their diagnosis and management of illness probably go some way toward effecting a cure. The same would hold for CAM therapists, perhaps more so. The memeplex[42] of certainty, overconfidence, autonomy, and an all-knowing paternalism have propagated extensively within the medical culture even though, as is sometimes the case with memes, it may not benefit clinician or patients over the long term.

Many variables are known to influence overconfidence behaviors, including ego bias,[43] gender,[44] self-serving attribution bias,[45] personality,[46] level of task difficulty,[47] feedback efficacy,[48,49] base rate,[30] predictability of outcome,[30] ambiguity of evidence,[30] and presumably others. A further complication is that some of these variables are known to interact with each other. It would be expected, too, that the level of critical thinking[1,50] would have an influence on overconfidence behavior. To date, the literature regarding medical decision making has paid relatively scant regard to the impact of these variables. Because scientific rigor ap-

**Table 3** Sources of overconfidence and strategies for correction

| Source | Correcting strategy |
| --- | --- |
| Lack of awareness and insight into decision theory | Introduce specific training in current decision theory approaches at the undergraduate level, emphasizing context dependency as well as particular vulnerabilities of different decision-making modes |
| Cognitive and affective bias | Specific training at the undergraduate level in the wide variety of known cognitive and affective biases. Create files of clinical examples illustrating each bias with appropriate correcting strategies |
| Limitations in feedback | Identify speed and reliability of feedback as a major requirement in all clinical domains, both locally and systemically |
| Biased evidence gathering | Promote adoption of cognitive forcing strategies to take account of disconfirming evidence, competing hypotheses, and consider-the-opposite strategy |
| Denial of uncertainty | Specific training to overcome personal and cultural barriers against admission of uncertainty, and acknowledgement that certainty is not always possible. Encourage use of "not yet diagnosed" |
| Base rate neglect | Make readily available current incidence and prevalence data for common diseases for particular clinical groups in specific geographical area |
| Context binding | Promote awareness of the impact of context on the decision-making process; advance metacognitive training to detach from the immediate pull of the situation and decontextualize the clinical problem |
| Limitations on transferability | Illustrate how biases work in a variety of clinical contexts. Adopt universal debiasing approaches with applicability across multiple clinical domains |
| Lack of critical thinking | Introduce courses early in the undergraduate curriculum that cover the basic principles of critical thinking, with iteration at higher levels of training |

pears lacking for System 1, the prevailing research emphasis in both medical[51] and other domains[52] has been on System 2.

## SOLUTIONS AND CONCLUSIONS

Overconfidence often occurs when determining a course of action and, accordingly, should be examined in the context of judgment and decision making. It appears to be influenced by a number of factors related to the individual as well as the task, some of which interact with one another. Overconfidence is associated in particular with confirmation bias and may underlie hindsight bias. It seems to be especially dependent on the manner in which the individual gathers evidence to support a belief. In medical decision making, overconfidence frequently is manifest in the context of delayed and missed diagnoses,[6] where it may exert its most harmful effects.

There are a variety of explanations why individual physicians exhibit overconfidence in their judgment. It is recognized as a common cognitive bias; additionally, it may be propagated as a component of a prevailing memeplex within the culture of medicine.

Numerous approaches may be taken to correct failures in reasoning and decision making.[2] Berner and Graber[6] outline the major strategies; **Table 3** expands on some of these and suggests specific corrective actions. Presently, no 1 strategy has demonstrated superiority over another, although, as noted earlier, several studies[38,39] suggest that when the generation of evidence is unbiased by giving competing hypotheses as much attention as the preferred hypothesis, overconfidence is reduced. Inevitably, the solution probably will require multiple paths.[16]

Prompt and reliable feedback about decision outcomes appears to be a prerequisite for calibrating clinician performance, yet it rarely exists in clinical practice.[41] From the standpoint of clinical reasoning, it is disconcerting that clinicians often are unaware of, or have little insight into, their thinking processes. As Epstein[53] observed of experienced clinicians, they are "less able to articulate what they do than others who observe them," or, if articulation were possible, it may amount to no more than a credible story about what they believe they might have been thinking, and no one (including the clinician) can ever be sure that the account was accurate.[16] But this is hardly surprising as it is a natural consequence of the dominance of System 1 thinking that emerges as one becomes an expert. As noted earlier, conscious practice of System 2 strategies can get compiled in TASS and eventually shape TASS responses. A problem once solved is not a problem; experts are expert in part precisely because they have solved most problems before and need only recognize and recall a previous solution. But this means that much of expert thinking is, and will remain, an invisible process. Often, the best we can do is make inferences about what thinking might have occurred in the light of events that subsequently transpired. It would be reassuring to think that with the development of expertise comes a reduction in overconfidence, but this is not always the case.[54]

Seemingly, clinicians would benefit from an understanding of the 2 types of reasoning, providing a greater awareness of the overall process and perhaps allowing them to explicate their decision making. Whereas System 1 thinking is unavailable to introspection, it is available to observation and metacognition. Such reflection might facilitate greater insight into the overall blend of decision-making modes typically used in the clinical setting.

Educational theorists in the critical thinking literature have expressed long-standing concerns about the need for introducing critical thinking skills into education. As van Gelder and colleagues[50] note, a certain level of competence in informal reasoning normally occurs through the processes of maturation, socialization, and education but few people actually progress beyond an everyday working level of performance to genuine proficiency.

This issue is especially relevant for medical training. The implicit assumption is made that by the time students have arrived at this tertiary level of education, they will have achieved appropriate levels of competence in critical thinking skills, but this is not necessarily so.[1] Though some will become highly proficient thinkers, the majority will probably not, and there is a need for the general level of reasoning expertise to be raised. In particular, we require education about detachment, overcoming belief bias effects, perspective switching, decontextualizing,[13] and a variety of other cognitive debiasing strategies.[55] It would be important, for example, to raise awareness of the many shortcomings and pitfalls of uncritical thinking at the medical undergraduate level and provide clinical cases to illustrate them. At a more general level, consideration should be given to introducing critical thinking training in the undergraduate curriculum so that many of the ~50 cognitive and affective biases in thinking[28] could be known and better understood.

Theoretically, it should be possible to improve clinical reasoning through specific training and thus reduce the prevalence of biases such as overconfidence; however, we should harbor no delusions about the complexity of the task. To reduce cognitive bias in clinical diagnosis requires far more than a brief session on cognitive debiasing. Instead, it is likely that successful educational strategies will require repeated practice and failure with feedback, so that limitations of transfer can be overcome. While some people have enjoyed success at demonstrating improved reasoning expertise with training,[30,56–59] to date there is little evidence that these skills can be applied to a clinical setting.[60] Nevertheless, it is a reasonable expectation that training in critical thinking,[1,61] and an understanding of the nature of cognitive[55] and affective bias,[62] as well as the informal logical fallacies that underlie poor reasoning,[28] would collectively lead to an overall improvement in decision making and a reduction in diagnostic failure.

Pat Croskerry, MD, PhD,
*Department of Emergency Medicine*
*Dalhousie University*
*Halifax, Nova Scotia, Canada*

Geoff Norman, PhD
*Department of Clinical Epidemiology and Biostatistics*
*McMaster University*
*Hamilton, Ontario, Canada*

## AUTHOR DISCLOSURES

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

**Pat Croskerry, MD, PhD,** has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

**Geoff Norman, PhD,** has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

## References

1. Jenicek M, Hitchcock DL. *Evidence-based practice: logic and critical thinking in medicine*. US: American Medical Association Press, 2005: 118–137.
2. Croskerry P. Timely recognition and diagnosis of illness. In: MacKinnon N, Nguyen T, eds. *Safe and Effective: the Eight Essential Elements of an Optimal Medication-use System.* Ottawa, Ontario: Canadian Pharmacists Association, 2007.
3. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study 1. *N Eng J Med.* 1991;324:370–376.
4. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian Health Care Study. *Med J Aust.* 1995;163:458–471.
5. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care.* 2000;38:261–271.
6. Berner E, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(suppl 5A):S2–S23.
7. Norman GR. The epistemology of clinical reasoning: perspectives from philosophy, psychology, and neuroscience. *Acad Med.* 2000; 75(suppl):S127–S133.
8. Bruner JS. *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press, 1986.
9. Hammond KR. Intuitive and analytic cognition: information models. In: Sage A, ed. *Concise Encyclopedia of Information Processing in Systems and Organizations.* Oxford: Pergamon Press, 1990:306–312.
10. Sloman SA. The empirical case for two systems of reasoning. *Psychol Bull.* 1996;119:3–22.
11. Evans J, Over D. *Rationality and Reasoning.* Hove, East Sussex, UK: Psychology Press, 1996.
12. Stanovich KE, West RF. Individual differences in reasoning: implications for the rationality debate? *Behav Brain Sci.* 2000;23:645–665; discussion 665–726.
13. Stanovich KE. *The Robot's Rebellion: Finding Meaning in the Age of Darwin.* Chicago: University of Chicago Press, 2005.
14. Zajonc RB. Feeling and thinking: preferences need no inferences. *Am Psychol.* 1980;35:151–175.
15. Fodor J. *The Modularity of Mind.* Cambridge, MA: MIT Press, 1983.

16. Norman G. Building on experience—the development of clinical reasoning. *New Engl J Med.* 2006;355:2251–2252.
17. Norman GR, Brooks LR. The non-analytic basis of clinical reasoning. *Adv Health Sci Educ Theory Pract.* 1997;2:173–184.
18. Hatala R, Norman GR, Brooks LR. Influence of a single example upon subsequent electrocardiogram interpretation. *Teach Learn Med.* 1999; 11:110–117.
19. Medin DL. Concepts and conceptual structure. *Am Psychol.* 1989;44: 1469–1481.
20. Brooks LR. Decentralized control of categorization: the role of prior processing episodes. In: Neisser U, ed. *Concepts and Conceptual Development*. Cambridge, UK: Cambridge University Press, 1987: 141–174.
21. Gigerenzer G, Todd P. *ABC Research Group. Simple Heuristics That Make Us Smart.* New York: Oxford University Press, 1999.
22. Eva KW, Norman GR. Heuristics and biases: a biased perspective on clinical reasoning. *Med Educ.* 2005;39:870–872.
23. Kulatunga-Moruzi C, Brooks LR, Norman GR. Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teach Learn Med.* 2001;13:110–116.
24. Ark TK, Brooks LR, Eva KW. Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Acad Med.* 2006;81:405–409.
25. Gladwell M. *Blink: The Power of Thinking Without Thinking.* New York: Little Brown and Co, 2005.
26. Dijksterhuis A, Bos MW, Nordgren LF, van Baaren RB. On making the right choice: the deliberation-without-attention effect. *Science.* 2006;311:1005–1007.
27. Zhaoping L, Guyader N. Interference with bottom-up feature detection by higher-level object recognition. *Curr Biol.* 2007;17:26–31.
28. Risen J, Gilovich T. Informal logical fallacies. In: Sternberg RJ, Roediger HL, Halpern DF, eds. *Critical Thinking in Psychology.* New York: Cambridge University Press, 2007.
29. Baron J. *Thinking and Deciding*.. 3rd ed. New York: Cambridge University Press, 2000.
30. Griffin D, Tversky A. The weighing of evidence and the determinants of confidence. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment.* New York: Cambridge University Press2002;230–249.
31. Yates JF, Lee JW, Sieck WR, Choi I, Price PC. Probability judgment across cultures. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment.* New York: Cambridge University Press, 2002:271–291.
32. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med.* 2005;80(suppl):S46–S54.
33. Krueger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999;77:1121–1134.
34. Haselton MG, Buss DM. Error management theory: a new perspective on biases in cross-sex mind reading. *J Pers Soc Psychol.* 2000;78:81–91.
35. Tiedens L, Linton S. Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing. *J Pers Soc Psychol.* 2001;81:973–988.
36. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol.* 1998;2:175–220.
37. Rozenblit LR, Keil FC, The misunderstood limits of folk science: an illusion of explanatory depth. *Cognit Sci.* 2002;26:521–562.
38. Koriat A, Lichtenstein S, Fischoff B. Reasons for confidence. *J Exp Psychol [Hum Learn].* 1980;6:107–118.
39. Arkes H, Faust D, Guilmette T, Hart K. Eliminating the hindsight bias. *J Appl Psychol.* 1988;73:305–307.
40. Katz J. Why doctors don't disclose uncertainty. In: Elstein A, Dowie J, eds. *Professional Judgment: A Reader in Clinical Decision Making.* Cambridge, UK: Cambridge University Press; 1988:544–565.
41. Croskerry PG. The feedback sanction. *Acad Emerg Med.* 2000;7: 1232–1238.
42. Blackmore S. *The Meme Machine.* Oxford: Oxford University Press, 1999.
43. Detmer DE, Fryback DG, Gassner K. Heuristics and biases in medical decision making. *J Med Educ.* 1978;53:682–683.
44. Lundeberg MA, Fox PW, Punccohar J. Highly confident but wrong: gender differences and similarities in confidence judgments. *J Educ Psychol.* 1994; 86:114–121.
45. Deaux K, Farris E. Attributing causes for one's own performance: the effects of sex, norms, and outcome. *J Res Pers.* 1977;11:59–72.
46. Landazabal MG. Psychopathological symptoms, social skills, and personality traits: a study with adolescents Spanish. *J Psychol.* 2006;9: 182–192.
47. Fischhoff B, Slovic P. A little learning. . .: Confidence in multicue judgment. In: Nickerson R, ed. *Attention and Performance.* VIII. Hillsdale, NJ: Erlbaum; 1980.
48. Lichenstein S, Fischoff B. Training for calibration. *Organ Behav Hum Perform.* 1980;26:149–171.
49. Arkes HR, Christensen C, Lai C, Blumer C. Two methods of reducing overconfidence. *Organ Behav Human Decis Process.* 1987;39:133–144.
50. van Gelder T, Bissett M, Cumming G. Cultivating expertise in informal reasoning. *Can J Exp Psychol.* 2004;58:142–152.
51. Croskerry P. The theory and practice of clinical decision making. *Can J Anaesth.* 2005; 52:R1–R8.
52. Dane E, Pratt MG. Exploring intuition and its role in managerial decision making. *Acad Manage Rev.* 2007;32:33–54.
53. Epstein RM. Mindful practice. *JAMA.* 1999;9:833–839.
54. Henrion M, Fischoff B. Assessing uncertainty in physical constants. *Am J Phys.* 1986; 54:791–798.
55. Croskerry P. The importance of cognitive errors in diagnosis and strategies to prevent them. *Acad Med.* 2003;78:1–6.
56. Krantz DH, Fong GT, Nisbett RE. Formal training improves the application of statistical heuristics in everyday problems. Murray Hill, NJ: Bell Laboratories, 1983.
57. Lehman DR, Lempert RO, Nisbett RE. The effects of graduate training on reasoning: formal discipline and thinking about everyday life events. *Am Psychol.* 1988;43:431–432.
58. Nesbett RE, Fong GT, Lehman DR, Cheng PW. Teaching reasoning. *Science.* 1987;238:625–631.
59. Chi MT, Glaser R, Farr MJ, *The Nature of Expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
60. Graber MA. Metacognitive training to reduce diagnostic errors: ready for prime time? [abstract]. *Acad Med.* 2003;78:781.
61. Lehman D, Nisbett R. A longitudinal study on the effects of undergraduate education on reasoning. *Dev Psychol.* 1990;26:952–960.
62. Croskerry P. The affective imperative: Coming to terms with our emotions. *Acad Emerg Med.* 2007;14:184–186.

The American Journal of Medicine (2008) Vol 121 (5A), S30–S33

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

# Expanding Perspectives on Misdiagnosis

A significant insight to emerge from the review of the diagnostic failure literature by Drs. Berner and Graber[1] is that the gaps in our knowledge far exceed the soundly established areas, particularly if we focus on empirical findings based on real-world work by real physicians. This lack of knowledge about the nature of diagnostic problems seems odd, given the current climate of concern and concentrated effort to address safety issues in healthcare, and especially given the centrality of diagnosis in the minds of practitioners. How is it that our knowledge about diagnosis—historically the most central aspect of clinical practice and one that directs the trajectory of tests, procedures, treatment choices, medications, and interventions—has been so impoverished?

## GAPS IN RESEARCH AND ANALYSIS

The knowledge gap does not appear to be due to lack of interest in how physicians arrive at a diagnosis. There has been considerable research aimed at identifying and describing the diagnostic process and the nature of diagnostic reasoning. However, the lack of progress in applying research findings to the messy world of clinical practice suggests that we might benefit from examination of an expanded set of questions. There are at least 5 areas in which a change of direction might lead to sustained progress.

### Diagnostic Models

A great deal of the work to date has assumed that diagnostic thinking is best described by highly rationalized analytic models of reasoning (e.g., the hypothetico-deductive or the Bayesian probabilistic models[2,3]), with little or no consideration of alternative approaches. There are some exceptions, including criticisms of this view (see Berg and colleagues[4,5] and Toulmin[6]), Norman's research on clinical reasoning,[7,8] and Patel and colleagues'[9] studies of medical decision making. Nevertheless, the prevailing view in healthcare continues to be that analytic models of reasoning describe optimal diagnostic process, i.e., that they are normative. If physicians are not employing these analytic processes, the assertion is that they ought to be.

Surprisingly, research in a number of complex fields has demonstrated that under conditions of uncertainty, time pressure, shifting and conflicting goals, high risk, and responsibility for dealing with multiple other actors in the situation, experts seldom engage in highly analytic modes of decision making. Rather, under these conditions, experts are most likely to use fast and generally sufficient strategies. These strategies (and the methods employed to study them) have been described within a research paradigm referred to as "naturalistic decision making."[10–13] These findings indicate that we need to better understand the full range of decision making and diagnostic strategies employed by physicians and the contexts of their use.

### Static Versus Dynamic Decision Problems

Most of the research performed regarding diagnosis in medical contexts has concerned static decision problems: only 1 decision needs to be made, the situation does not change, and the alternatives are clear. (A typical example is deciding whether a radiograph contains a fracture). However, much of the work of medicine concerns dynamic decision problems: (1) a series of interdependent decisions and/or actions is required to reach the goal; (2) the situation changes over time, sometimes very rapidly; (3) goals shift or are redefined. Decisions that the clinician make change the milieu, resulting in a new challenge to resolve.[14] In contrast to static problems, in dynamic problems there is no theory or process element even close to being considered normative, either for approaching the problem or for establishing a particular sequence of decisions and/or actions as correct.

### Problem Detection and Recognition

One of the greatest holes in our current knowledge base is the failure to address issues of problem detection and recognition. Diagnostic problems do not present themselves fully formed like pebbles lying on a beach. The

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

Requests for reprints should be addressed to Beth Crandall, Klein Associates Division, Applied Research Associates, 1750 Commerce Center Boulevard North, Fairborn, Ohio 45324-6362.

E-mail address: bcrandall@decisionmaking.com.


doi:10.1016/j.amjmed.2008.02.002

understanding that an event represents a "problem" must instead be constructed from circumstances that are puzzling, troubling, uncertain, and possibly irrelevant. In order to discern the problem contained within a particular set of circumstances, practitioners must make sense of an uncertain and disorganized set of conditions that initially make little sense.[15,16] Here, much of the work of diagnosis consists of preconscious acts of perception[10,17–19] and sense making by clinicians who use a variety of strategies to discern the real-world context.[13] Given a stream of passing phenomena, distinguishing between items that are relevant or irrelevant, and those that must be accounted for compared with those that can be discounted, creates a preconscious framing that bounds the problem of diagnosis before it is ever consciously considered. This is an important task that has been inadequately studied. If we are going to understand how problems are missed or misunderstood, we need to understand the processes involved in their detection and recognition.

### Centrality

Traditionally, diagnosis has been considered medicine's central task, but it might be useful to entertain the possibility that this emphasis may be misdirected. Having a solid diagnosis often makes much of clinical work easier. However, the lack of a firm diagnosis does not relieve the practitioner of the necessity to take action, and by taking action, risk that the world will be changed, perhaps in unintended ways. Thus, one might argue that the central task of medicine is not diagnosis, but management, especially management in the face of uncertainty. Stated another way, the central question of clinical work might not be, "What is the diagnosis?" but rather, "What should we do now?"

### Individual Versus Distributed Cognition

Most research on diagnostic decision making has concentrated almost entirely on what goes on inside physicians' minds, focusing on internal mental processes, including various cognitive biases and simplifying heuristics. Although understanding the individual physician's cognitive work is clearly necessary, it is not sufficient. Clinicians do their work while embedded in a complex milieu of people, artifacts, procedures, and organizations. All these factors can contribute or detract from diagnostic performance in complex ways; the possibility that the diagnostic process may go awry for reasons other than the physician's reasoning abilities needs more attention. Considering physicians and their environment as joint cognitive systems,[20] where cognition and expertise are distributed across multiple people, objects, and procedures within a clinical setting,[21] offers a way to widen the tight focus from "inside the physician's head" so that we can begin to examine this larger, and far more complex, scenario.

## COMPLEXITIES SURROUNDING DIAGNOSIS

One reason we know so little about diagnostic problems may be the complexity of the systems and work processes that surround diagnosis. We know that differences in diagnostic performances exist, but we do not understand diagnostic failure in any deep or detailed way. In the emergency department, for example, the physician's diagnostic process is carried out within the context of large numbers of patients, many of whom have multiple problems; there is little time, resources are constrained, and conditions are chaotic. Some possibilities worth considering include:

- **Context:** In what situations, and under what conditions, are diagnostic failures most and least prevalent? We need to understand the real-world contexts in which medical diagnosis occurs.
- **Team influences:** The individual physician is surrounded by other healthcare providers, including other clinicians, who share responsibility for patient care and outcome. How does the distributed nature of patient care foster or prevent diagnostic failure? In the field of aviation, implementation of crew resource management (CRM) has been credited with significant improvements in aviation safety. CRM requires that the pilot in the second seat voice concern to the captain and take assertive action if those matters are ignored. Is aviation's example a useful analogue? In what ways is it applicable?
- **System influences:** Some hospital systems have been highly successful in addressing patient safety issues such as medication errors and nosocomial infections. Presumably, the prevalence and severity of diagnostic failure vary considerably among hospital systems. This leads to the question, What system-level practices foster diagnostic quality?
- **Individual differences:** All physicians make mistakes but they appear to occur more frequently among some practitioners, even within a given specialty.[22,23] We know that with experience, diagnostic performance improves but that such progress is not invariant. Some physicians become extraordinarily skilled at evaluation and are recognized by their peers as the "go to" person for the toughest diagnostic challenges. Understanding the elements leading to such expertise would surely be informative, as would gleaning why experience appears to enhance the diagnostic performance of some physicians more than others.

## DESIGNING EFFECTIVE FEEDBACK MECHANISMS

Identifying the sources of diagnostic failure is a critical first step towards creating feedback systems that provide leverage on the problem. Finding ways to provide feedback on diagnostic performance seems an important venue for improvement, however many difficulties exist. Thus, simply providing feedback is not a "magic bullet" automatically leading to improvement. Learning specialists have found that feedback has greatest impact when it is specific, detailed, and timely.[24] These 3 issues, and a 4th—the differential values assigned to different types of failure—represent significant challenges to designing effective feedback systems for physicians.

### Specificity

Providing overall data about diagnostic error rates in physicians is unlikely to get us very far. Grouped data and general findings leave too much room for individual physicians to distance themselves from the findings. However, the processes by which individual physicians' diagnostic performance might be tracked, tagged, and reported back to them are not immediately apparent or readily available.

### Detail

To be effective, feedback must give physicians information that illuminates contingent relationships and causal sequences. Otherwise, they are left with unhelpful admonitions such as "work harder, don't make mistakes, maintain a high index of suspicion." Feedback needs to provide clinicians with sufficient information so that they can move in an adaptive direction. The simpler the system, the more helpful statistical quality control data are as a basis for self-correction. Highly complex systems may prove insufficient because they create dense forests of information that people—even highly educated, experienced people—have a great deal of difficulty navigating. More data are not necessarily helpful. In many cases, people do not need more data; they need help in making meaning of the data they have.

### Timeliness

The timeliness of feedback, especially regarding diagnostic performance, may be particularly problematic, as the "final diagnosis" often is not known for some time and, indeed, sometimes is never known. Furthermore, in some settings, delayed feedback can disastrously worsen, rather than improve, performance.[14]

### Differential Value

Finally, simple feedback mechanisms may lead physicians to become systematically inaccurate in undesirable ways, owing to differences in value ascribed to various types of failures. For example, feedback to an emergency physician showing that he/she discharged a patient who subsequently proved to have an acute myocardial infarction is likely to have a much different impact on behavior than feedback showing that a patient admitted for chest pain proved not to have an acute coronary syndrome. The former is likely to be viewed as an adverse event with a significant affective impact while the latter may be perceived as a nonevent.

## CONCLUSION

Diagnostic failures are both manifestly important and difficult to comprehend in useful ways. We need to provide a rich fabric of information that allows members of the medical community to see what works and what does not, to hone diagnostic skill, and to hold one another accountable for the quality of diagnoses. To do this, we need to enlarge our notions of the nature of clinical work and of human performance in complex, conflicted, and uncertain contexts.

Beth Crandall, BS
*Klein Associates Division*
*Applied Research Associates*
*Fairborn, Ohio, USA*

Robert L. Wears, MD, MS
*Department of Emergency Medicine*
*University of Florida Health Science Center*
*Jacksonville, Florida, USA*

## AUTHOR DISCLOSURES

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

**Beth Crandall, BS,** has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

**Robert L. Wears, MD, MS,** has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

## References

1. Berner E, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(suppl x):xx–xx.
2. Kassirer JP. Diagnostic reasoning. *Ann Intern Med.* 1989;110:893–900.
3. McNeil BJ, Keller E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med.* 1975;293:211–215.
4. Berg M. *Rationalizing Medical Work.* Cambridge, MA: MIT Press, 1997.
5. Timmermans S, Berg M. *The Gold Standard: the Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia, PA: Temple University Press, 2003.
6. Toulmin S. *Return to Reason.* Cambridge, MA: Harvard University Press, 2001.
7. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ.* 2005;39:418–427.
8. Norman G. Building on experience—the development of clinical reasoning. *N Engl J Med.* 2006;355:2251–2252.
9. Patel VL, Kaufman DR, Arocha JF. Emerging paradigms of cognition in medical decision-making. *J Biomed Inform.* 2002;35:52–75.
10. Orasanu J, Connolly T. The reinvention of decision making. In: Klein GA, Orasanu J, Calderwood R, Zsambok CE, eds. *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Company, 1993.
11. Rasmussen J. Deciding and doing: decision making in natural contexts. In: Klein G, Orasanu J, Calderwood R, Zsambok CE, eds. *Decision Making in Action: Models and Methods.* Norwood, NJ: Ablex Publishing Company; 1993:158–171.
12. Klein G. *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press, 1998.
13. Rasmussen J. Diagnostic reasoning in action. *IEEE Transactions on Systems, Man and Cybernetics, Part A.* 1993;23:981–992.

14. Brehmer B. Development of mental models for decision in technological systems. In: Klein G, Orasanu J, Calderwood R, Zsambok CE, eds. *Decision Making in Action: Models and Methods.* Norwood, NJ: Ablex Publishing Company; 1993:111–120.
15. Weick KE. *Sensemaking in Organizations.* Thousand Oaks, CA: Sage Publications, Inc, 1995.
16. Wears RL, Nemeth CP. Replacing hindsight with insight: towards a better understanding of diagnostic failures. *Ann Emerg Med.* 2007;49: 206–209.
17. Klein G, Pliske R, Crandall B, Woods DD. Problem detection. *Cogn Technol Work.* 2005;7:14–28.
18. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ Theory Pract.* 1997;2:173–184.
19. Norman G, Brooks LR. The role of experience in clinical education. *Med Educ.* 2007; (in press).
20. Woods DD, Hollnagel E. *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering.* Boca Raton, FL: CRC Press/Taylor & Francis Group, 2006.
21. Hutchins E. *Cognition in the Wild.* Cambridge, MA: MIT Press, 1996.
22. Elstein AS, Schwarz A. Evidence base of clinical diagnosis: clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ.* 2002;324:729–732.
23. Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. *Cognit Sci.* 1986;10:91–116.
24. Glaser R, Bassock M. Learning theory and the study of instruction. *Annual Review of Psychology.* 1989; 40:631–636.

The American Journal of Medicine (2008) Vol 121 (5A), S34–S37

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

# Sidestepping Superstitious Learning, Ambiguity, and Other Roadblocks: A Feedback Model of Diagnostic Problem Solving

A central argument of Drs. Eta S. Berner and Mark L. Graber's review[1] is that feedback processes are crucial to enhancing or inhibiting the quality of diagnostic problem solving over time. Our goal is to enrich the conversation about diagnostic problem solving by presenting an explicit model of the feedback processes inherent in improving diagnostic problem solving. We present a simple, generic model of the fundamental feedback processes at play in calibrating or improving diagnostic problem-solving skill over time. To amplify these key processes, this commentary draws on a 50-year evidence and theory base from the discipline of system dynamics.[2,3]

Using Berner and Graber's analysis[1] of the challenges of feedback and calibration as a starting point, we depict how feedback loops can operate in a robust or benign manner to support and improve immediate and long-term diagnostic problem solving. Drawing on insights from research on how people manage problem solving that involves dynamic feedback, we then describe how this process is likely to break down. Finally, leverage points for improving diagnostic problem solving and avoiding error are provided.

To improve diagnostic problem solving, practitioners and researchers need to move away from viewing diagnosis as a "one-shot deal." When diagnosis is perceived as a stand-alone, discrete episode of judgment, the solutions suggested to resolve error focus on reducing cognitive biases and increasing expertise and vigilance at the individual clinician level. It is not that such recommendations have no merit, but simply that they are only a small piece of a much larger repertoire of possible solutions that come into sight when we regard diagnostic problem solving as a recursive, feedback-driven process. Put differently, rather than viewing diagnosis as an event or episode, we suggest emphasizing it as an active, ongoing practice in which clinicians revise and redraft their conclusions over time.[4-6]

## WHEN CALIBRATION WORKS: AN OPTIMAL FEEDBACK PROCESS

From the moment a clinician begins a patient encounter, he/she is selecting, labeling, and processing information (e.g., symptoms, results from studies, and other data) from the client or his record. The practitioner shapes this information into a diagnosis that, in turn, influences his/her view and collection of subsequent information. Discrete decisions made without feedback have been likened to hitting a target from a distance in one try; in contrast, diagnostic problem solving is analogous to a situation where one can monitor and correct the trajectory based on feedback.[4,5]

Patient care is a feedback process in which the clinician makes judgments and takes actions with the intended rationale of bringing the patient closer to the desired, presumably healthier, status. This process of observing/diagnosing/treating/observing describes a balancing or goal-seeking feedback loop, in which feedback about the patient's status allows a clinician to calibrate therapy over the very short term. Although physicians may be able to adjust a diagnosis and treatment based on conversation and examination during a specific patient encounter, Berner and Graber[1] argue that lack of timely or consistent feedback on the accuracy and quality of diagnoses over the long term makes it difficult for them to improve their diagnostic problem-solving skills over time. Once out of medical school and residency, most physicians operate in a "no news is good news" mode, believing that unless they hear about problems, the diagnoses they have made are correct. Berner and Graber invoke a well-established fact of learning theory, namely, that improvement is nearly impossible without accurate and timely feedback. Improving one's diagnostic problem-solving skill, they argue, requires an ability to calibrate the match between the diagnosis made and the patient's actual long-term status.

The generic feedback process that would allow a clinician to calibrate and improve a key element of long-term

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

Requests for reprints should be addressed to Jenny W. Rudolph, PhD, Center for Medical Simulation, 65 Lansdowne Street, Cambridge, Massachusetts 02139.

E-mail address: JWRudolph@partners.org.


doi:10.1016/j.amjmed.2008.02.003

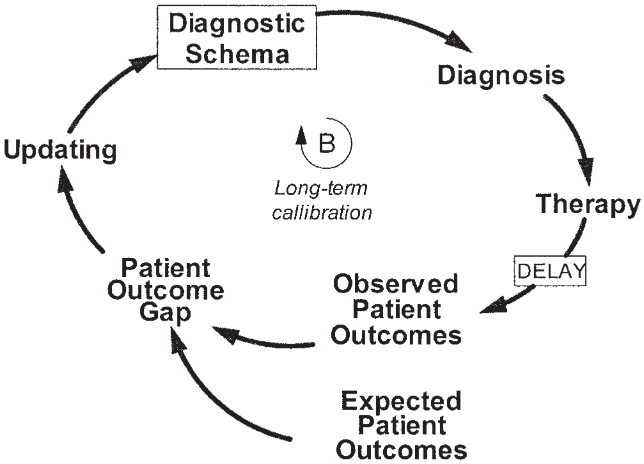


**Figure 1** Calibrating or improving diagnostic quality over time. The "B" labeled "long-term calibration" signifies a balancing loop that updates clinicians' diagnostic schema based on information that allows them to compare how they expect the patient to progress with the patient's observed outcomes. Arrows indicate the direction of causality.

diagnostic skill, the quality of his/her "diagnostic schemas," is depicted in **Figure 1**. A diagnosis is the result of applying a diagnostic schema to information about the patient as the clinician perceives it. *Schema* is a term from cognitive science referring to a person's mental model, or internal image of a given professional domain or area.[7] Schemas form the basis of processes such as "recognition-primed decision making" that allow clinicians to match a library of images of past experiences with the present constellation of signs and symptoms to formulate a diagnosis.[8]

The long-term feedback process in diagnosing and treating an individual patient depicted in Figure 1, like the short-term feedback process, is a balancing or adaptive process. It is a longer-term process of learning from experience, in which the clinician adjusts the diagnostic schema for the patient by comparing expected outcomes with observed actual outcomes. To illustrate how this loop operates, we start with *Diagnosis*. In making a *Diagnosis,* the clinician employs the current *Diagnostic Schema*, developed through training and experience, to interpret patient information and recommend a specific course of *Therapy*. Based on the the therapy recommended, the clinician expects the patient's condition will evolve in a certain way to yield *Expected Patient Outcomes*. Ideally, after some time has elapsed for the therapy to take effect, the clinician sees the actual *Observed Patient Outcomes*. Comparing the *Observed Patient Outcomes* with *Expected Patient Outcomes* (this comparison is often tacit or unconscious), the clinician then identifies the *Patient Outcome Gap*, which stimulates *Updating* or revising of the existing *Diagnostic Schema*. In optimal settings, this schema accounts well for the patient's history, constellation of signs and symptoms, and treatment results. To the extent that the diagnostic schema improves, the quality of the clinican's diagnoses at later patient encounters also improves.

## BARRIERS TO IMPROVING DIAGNOSTIC PROBLEM-SOLVING SKILL OVER TIME

Berner and Graber[1] and other contributors to this supplement note that a simple but significant barrier to enhancing diagnostic problem-solving skill over time is that the link between therapy and observed patient outcomes often is nonexistent. In the absence of significant information provided by autopsy, data from downstream clinicians, or tailored quality measures, clinicians are unable to update their diagnostic schema. Several decades of research on how people manage information in the face of dynamic feedback reveal other challenges as well. We highlight 3 significant barriers to updating diagnostic schema in a sound way: delays, ambiguous feedback, and superstitious learning.[2,9,10]

### Delays

For both an immediate patient encounter and the long-term process of improving and updating one's diagnostic schema, delays in feedback can cause problems. Delays slow the accumulation of evidence and create fluctuations in evidence that make it difficult to draw sound conclusions.[9] Obviously, as the length of time between therapy and its impact increases, the likelihood that the physician will observe the outcome decreases. Examples of this include patients who do not experience the full consequences of the therapy or physicians who do not see the patient again, thereby rendering outcome feedback unavailable. Time delays, thus, partially explain why the link from therapy to observed patient outcomes may be so weak, as Berner and Graber[1] suggest.

Delays compromise learning even when outcome feedback is available. Delays between cause and effect make inferences about causality far more difficult because they give rise to a characteristic of feedback systems known as *dynamic complexity*.[2,9] In diagnostic problem solving, dynamic complexity can take the form of unexpected oscillations between desired and undesired therapeutic outcomes, amplification of certainty on the part of the clinician (e.g., fixation), and excessive or diminished commitment to particular treatments.[11] For example, if effects from therapy occur after the physician's felt need to move forward with patient care, he/she may pursue contraindicated interventions or drop indicated ones—continuing to intervene although curative measures have been taken or failing to intervene although treatment has been inadequate. Research repeatedly has demonstrated the failure to learn in situations with even modest amounts of dynamic complexity.[9] Finally, time delays quite simply slow down the completion of the feedback loop; longer delays mean fewer learning cycles in any time period.

### Ambigious Feedback

Although a clinican may receive feedback about how his/her diagnosis and therapy has influenced the patient, effectiveness can be compromised because such feedback often is ambiguous. The primary problem is that changes in the patient's observed status caused by the physician's actions

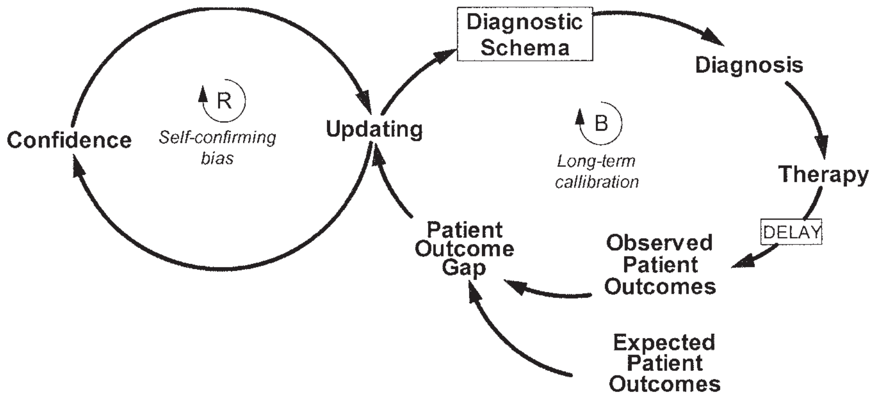


**Figure 2** How confidence impedes calibration.The "R" labeled "self-confirming bias" signifies a reinforcing loop that amplifies clinicians' confidence in their current diagnostic problem-solving skill. Arrows indicate the direction of causality.

are influenced by a range of other clinical and lifestyle variables both inside and outside the clinician's control. Confusingly, data about their patients can equally support a wide variety of clinical conclusions, making it difficult for physicians to assess what actions actually work best. Controlled experimentation is almost never possible in real clinical settings. Ambiguous information invites subjective interpretation, and, like many people, physicians tend to make self-fulfilling interpretations (e.g., "The diagnosis was correct") in the face of such ambiguity, perhaps missing the opportunity to update flawed diagnostic schema.

### Superstitious Learning

In the face of time delays and ambiguity, superstitious learning thrives. Sterman[9] relates the case of Baseball Hall of Fame hitter Wade Boggs, who ate chicken every game day for years because he had played well once following a dinner of lemon chicken. While this might seem laughable, ambiguous or weak feedback supports "strong but wrong" self-confirming attributions about what works.[12,13] During the time gap between therapy and observed outcome, much transpires that the clinician does not directly observe. Physicians, like other people, fill in the blanks with their own superstitious explanations—conclusions that fit the data but are based on weak or spurious correlations (e.g., eating chicken improves baseball performance).

The lessons of superstitious learning persist because satisfactory explanations (e.g., scurvy is an unavoidable result of lengthy sea voyages) suppress the search for better answers (e.g., scurvy results from vitamin C deficiency). Recent studies show that only about 15% of physicians' decisions are evidence based; weak or ambiguous feedback contributes to this situation by preventing physicians from learning when their self-confirming routines are inappropriate, inaccurate, or dangerous.

## HOW CONFIDENCE CAN DISRUPT LEARNING

How does such pseudolearning persist? Berner and Graber[1] argue that confidence or overconfidence plays a role. The feedback process we have described (Figure 1) is a balancing loop that attempts to close the gap between expected and observed patient outcomes. When that gap does not close, clinicians should seek additional or alternative data. But Berner and Graber show that often does not happen. To understand why, we introduce another feedback loop in **Figure 2**.

To understand the impact of the self-confirming bias loop (Figure 2), the contrast between the process by which physicians ideally update their diagnostic schema and the actual one described by Berner and Graber[1] should be kept in mind: In the adaptive scenario, where learning occurs when *Therapy* influences the *Observed Patient Outcomes*, the physician observes these outcomes and is informed by the *Patient Outcome Gap*. In situations where the link between *Therapy* and *Observed Patient Outcomes* is nonexistent or weak, the *Patient Outcome Gap* is either unknown or unclear.

Berner and Graber[1] argue that in the absence of such clear feedback, physicians feel little need to update their current *Diagnostic Schema*. Thus, a felt need for *Updating* declines and *Confidence* increases. As *Confidence* increases, the felt need for *Updating* decreases further in a reinforcing cycle. While calibrating or improving one's diagnostic problem solving already faces the significant challenges posed by missing or ambiguous feedback, lack of feedback also triggers a vicious reinforcing cycle that erroneously amplifies confidence. It is this reinforcing confidence cycle that is the nail in the coffin of robust learning that would allow clinicians to improve diagnostic problem solving over time.

In conclusion, we ask, "Does a doctor who has practiced for 30 years have a lower rate of diagnostic error than a doctor who has practiced for 5 years?" If the feedback processes we have described were functioning optimally, the answer should be a resounding "Yes!" Based on the review by Berner and Graber,[1] however, the answer is unclear. To contribute to policies that reduce the rate of diagnostic errors, we have highlighted 2 faces of the balancing feedback processes that drive diagnostic problem solving. These processes can function adaptively, improving diagnostic schema over time and problem solving during a patient encounter. If physicians in practice for 30 years had a notably lower rate of diagnostic error than their rookie counterparts, it would indicate these loops were functioning well. But these processes break down when crucial links are weakened or do not function at all. When this happens, adaptive learning processes are further hobbled by a vicious reinforcing cycle that maintains or amplifies a misplaced sense of confidence.

If, as scholars of human judgment have argued, overconfidence is a highly ingrained human trait, trying to reduce it is a Sisyphean task.[14] The leverage points for this uphill task lie, as our colleagues in this supplement have argued, in systematically assuring that downstream feedback is (1) available and (2) as unambiguous as possible so that physicians experience a felt need to update their diagnostic schema. It is this pressure to update that can weaken the reinforcing confidence loop.

Jenny W. Rudolph, PhD
*Center for Medical Simulation*
*Cambridge, Massachusetts, USA*
*Harvard Medical School*
*Cambridge, Massachusetts, USA*

J. Bradley Morrison, PhD
*Brandeis University International Business School*
*Waltham, Massachusetts, USA*

## AUTHOR DISCLOSURES

The authors report the following conflicts of interest with the sponsor of this supplement article or products discussed in this article.

**Jenny W. Rudolph, PhD,** has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

**J. Bradley Morrison, PhD,** has no financial arrangement or affiliation with a corporate organization or a manufacturer of a product discussed in this article.

## References

1. Berner E, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(suppl 5A):S2–S23.
2. Forrester JW. *Industrial Dynamics.* Portland, OR: Productivity Press, 1994.
3. Sterman J. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Homewood, IL: Irwin/McGraw Hill, 2000.
4. Hogarth RM. Beyond discrete biases: functional and dysfunctional aspects of judgmental heuristics. *Psychol Bull.* 1981;90:197–217.
5. Kleinmuntz DN. Cognitive heuristics and feedback in a dynamic decision environment. *Manage Sci.* 1985;31:680–702.
6. Weick KE, Sutcliffe K, Obstfeld D. Organizing and the process of sensemaking. *Organ Sci.* 2005;16:409–421.
7. Gentner D, Stevens AL. *Mental Models.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.
8. Klein G. *Sources of Power: How People Make Decisions.* Cambridge, MA: MIT Press, 1998.
9. Sterman JD. Learning from evidence in a complex world. *Am J Public Health.* 2006;96:505–514.
10. Repenning NP, Sterman JD. Capability traps and self-confirming attribution errors in the dynamics of process improvement. *Adm Sci Q.* 2002;47:265–295.
11. Rudolph JW, Morrison JB. Confidence, error and ingenuity in diagnostic problem solving: clarifying the role of exploration and exploitation. Paper presented at: Annual Meeting of the Academy of Management; August 5–8, 2007; Philadelphia, PA.
12. Sterman J, Repenning N, Kofman F. Unanticipated side effects of successful quality programs: exploring a paradox of organizational improvement. *Manage Sci.* 1997;43:503–521.
13. Reason J. *Human Error.* New York: Cambridge University Press, 1990.
14. Bazerman MH. *Judgment in Managerial Decision Making*. New York: John Wiley and Sons, 1986.

The American Journal of Medicine (2008) Vol 121 (5A), S38–S42

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

# 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 Graber[3] 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: gschiff@partners.org.


doi:10.1016/j.amjmed.2008.02.004

**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, 9$^{th}$ Revision (ICD-9) or

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

| |
|---|
| • Diagnosis of severity/acuity<br>—Failure to recognize patient need to be hospitalized or sent to ICU |
| • Diagnosis of complication<br>—Assessing sequelae of a disease, drug, or surgery |
| • Diagnosis of a recurrence<br>—What follow-up surveillance is required and how to interpret results |
| • Diagnosis of cure or failure to respond<br>—When can clinician feel secure vs worry if symptoms don't improve<br>—When should "test-of-cure" be done routinely |
| • Diagnosis of a misdiagnosis<br>—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<br>— When does a clinician decide a patient is/is not responding |
| • Interpretation of partial responses |
| • How to incorporate known variations in response<br>—Timing<br>—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<br>—When and how often to repeat an x-ray or blood test |
| • Role of mitigating factors<br>—Self-limited illnesses<br>—Placebo response<br>—Naturally relapsing and remitting courses of disorders |

*Diagnostic and Statistical Manual of Mental Disorders*, 4$^{th}$ 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 physician 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 a 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 Najarian[14] 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 Graber[3] 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/advnaces/. 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.

The American Journal of Medicine (2008) Vol 121 (5A), S43–S46

ELSEVIER

THE AMERICAN JOURNAL *of* MEDICINE®

# Taking Steps Towards a Safer Future: Measures to Promote Timely and Accurate Medical Diagnosis

The issue of diagnostic error is just emerging as a major problem in regard to patient safety, although diagnostic errors have existed since the beginnings of medicine, millennia ago. From the historical perspective, there is substantial good news: medical diagnosis is more accurate and timely than ever. Advances in the medical sciences enable us to recognize and diagnose new diseases. Innovation in the imaging and laboratory sciences provides reliable new tests to identify these entities and distinguish one from another. New technology gives us the power to find and use information for the good of the patient. It is perfectly appropriate to marvel at these accomplishments and be thankful for the miracles of medical science.

It is equally appropriate, however, to take a step back and consider whether we are really where we would like to be in regard to medical diagnosis. There has never been an organized discussion of what the goal should be in terms of diagnostic accuracy or timeliness and no established process is in place to track how medicine performs in this regard. In the history of medicine, progress toward improving medical diagnosis seems to have been mostly a passive haphazard affair.

The time has come to address these issues. Every day and in every country, patients are diagnosed with conditions they don't have or their true condition is missed. Furthermore, patients are subjected to tests they don't need; alternatively, tests they do need are not ordered or their test reports are lost. Despite our best intentions to make diagnosis accurate and timely, we don't always succeed.

Our medical profession needs to consider how we can improve the accuracy and timeliness of diagnosis. Goals should be set, performance should be monitored, and progress expected. But where and how should this process be started? The authors in this supplement to *The American Journal of Medicine* focus on the physician's role in diagnostic error; a variety of strategies are offered to improve diagnostic calibration and reduce diagnostic errors. Although many of these strategies show potential, the pathway to accomplish their goals is not clear. In some areas, little research has been done while in others the results are mixed. We don't have easy ways to track diagnostic errors; no organizations are ready or interested to compile the data even if we did. Moreover, we are uncertain how to spark improvements and align motivations to ensure progress. Although our review[1] focuses on overconfidence as a pivotal issue in an effort to engage providers to participate in error-reducing strategies, this is just one suggestion among many; a host of other factors, both cognitive and system related, contribute to diagnostic errors.

For all of these reasons, a broader horizon is appropriate to address diagnostic error. My goal in this commentary is to survey a range of approaches with the hope of stimulating discussion about their feasibility and likelihood of success. This requires identifying all of the stakeholders interested in diagnostic errors. Besides the physician, who obviously is at the center of the issue, many other entities potentially influence the rate of diagnostic error. Foremost amongst these are healthcare organizations, which bear a clear responsibility for ensuring accurate and timely diagnosis. It is doubtful, however, that physicians and their healthcare organizations alone can succeed in addressing this problem.

At least in the short term then, we clinicians seek to enlist the help of another key stakeholder—the patient, who is typically regarded as a passive player or victim. Patients are in fact much more than that. Finally, there are clear roles that funding agencies, patient safety organizations, oversight groups, and the media can play to assist in the overall goal of error reduction. What follows is advice for each of these parties, based on our current—albeit incomplete and untested—understanding of diagnostic error (**Table 1**).

## HEALTHCARE SYSTEMS

Leaders of healthcare systems recognize the critical role their organizations play in promoting quality care and patient safety. Unfortunately, in the eyes of organization leaders, "patient safety" typically refers to injuries from falls, nosocomial infections, the "never" events, and medication

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

This research was supported by a grant from the National Patient Safety Foundation.

Requests for reprints should be addressed to Mark Graber, MD, Medical Service–III, Veterans Affairs Medical Center, Northport, New York 11768.

E-mail address: mark.graber@va.gov.


doi:10.1016/j.amjmed.2008.02.006

errors. Healthcare leaders need to expand their concept of patient safety to include responsibility for diagnostic errors, an area they traditionally have been happy to relegate to their physicians. Surprisingly, most diagnostic errors in medicine involve factors related to the healthcare system.[2] Addressing these problems could substantially reduce the likelihood of similar errors in the future. Even the cognitive aspects of diagnostic error can to some extent be mitigated by interventions at the system level. Leaders of healthcare organizations should consider these steps to help reduce diagnostic error.

### System-related Suggestions

**Ensure That Diagnostic Tests Are Done on a Timely Basis and That Results Are Communicated to Providers and Patients.** Insist that tests and procedures are scheduled and performed on a timely basis.[3] Monitor the turn around time of key tests, such as x-rays. Ensure that providers receive test results and that a surrogate system exists for providers who are unavailable. Unless this system functions flawlessly, establish a pathway for patients to receive critical test abnormalities directly, as a backup measure.

**Optimize Coordination of Care and Communication.** Develop electronic medical records so that patient data is available to all providers in all settings. Encourage interpersonal communication among staff via telephone, e-mail, and instant messaging. Develop formal and universal ways to communicate information verbally and electronically across all sites of care.

**Continuously Improve the Culture of Safety.** Include diagnostic errors as a routine part of quality assurance surveillance and review; identify any adverse events that appear repeatedly as possible examples of normalization of deviance. Monitor consultation timeliness. Ensure medical records are consistently available and reviewed. Strive to make diagnostic services available on weekend/night/holiday shifts. Minimize distractions and production pressures so that staff have enough time to think about what they are doing. Minimize errors related to sleep deprivation by attention to work hour limits, and allowing staff naps if needed.

### Suggestions Regarding Cognitive Aspects of Diagnosis

**Facilitate Perceptual Tasks.** Take advantage of suggestions from the human-factors literature on how to improve the detection of abnormal results. For example, graphic displays that show trends make it more likely that clinicians will detect abnormalities compared with single reports or tabulated lists; use of these tools could allow more timely appreciation of such matters as falling hematocrits or progressively rising prostate-specific antigen values. Computer-aided perception might help reduce diagnostic errors (e.g., as adjunct with mammograms to detect breast cancer). Controlled trials have shown that use of a computer algorithm can improve both the specificity and sensitivity of cancer detection more than an independent reading by a second radiologist.[4]

**Provide Tools for Decision Support.** Provide physicians with access at the point of care to the Internet, electronic medical reference texts and journals, and electronic decision-support tools. These resources have substantial potential to improve clinical decision making,[5] and their impact will increase as they become more accessible, more sophisticated, and better integrated into the everyday process of care.

**Have Appropriate Clinical Expertise Available When It's Needed.** Don't allow front-line clinicians to read and interpret x-rays. Ensure that all trauma patients are seen by a surgeon. Facilitate referral to appropriate subspecialists. Ensure that trainees are appropriately supervised. Encourage second readings for key diagnostic studies (e.g., Pap smears, anatomic pathology material that is possibly malignant) and encourage second opinions in general.

**Enhance Feedback to Improve Physician Calibration.** Encourage discussion of diagnostic errors. Encourage and reward autopsies and "morbidity and mortality" conferences; provide access to electronic counterparts, such as "Morbidity and Mortality (M & M) Rounds on the Web" sponsored by the Agency for Healthcare Research and Quality (AHRQ).[6] Establish pathways for physicians who saw the patient earlier to learn that the diagnosis has changed.

## PATIENTS

Patients obviously have the appropriate motivation to help reduce diagnostic errors. They are perfectly positioned to prevent, detect, and mollify many system-based as well as cognitive factors that detract from timely and accurate diagnosis. Properly educated, patients are ideal partners to help reduce the likelihood of error. For patients to act effectively in this capacity, however, requires that physicians orient them appropriately and reformulate, to some extent, certain aspects of the traditional relationship between themselves and their patients. Two new roles for patients to help reduce the chances for diagnostic error are proposed below.

### Be Watchdogs for Cognitive Errors

Traditionally, physicians share their initial impressions with a new patient, but only to a limited extent. Sometimes the suspected diagnosis isn't explicitly mentioned, and the patient is simply told what tests to have done or what treatment will be used. Patients could serve an effective role in checking for cognitive errors if they were given more information, including explicit disclosure of their diagnosis, its probability, and instructions on what to expect if this is correct. They should be told what to watch for in the

**Table 1** Recommendations to reduce diagnostic errors in medicine: stakeholders and their roles

| **Direct and Major** | Role |
|---|---|
| Physicians | • Improve clinical reasoning skills and metacognition<br>• Practice reflectively and insist on feedback to improve calibration<br>• Use your team and consultants, but avoid groupthink<br>• Encourage second opinions<br>• Avoid system flaws that contribute to error<br>• Involve the patient and insist on follow-up<br>• Specialize<br>• Take advantage of decison-support resources |
| Healthcare organizations | • Promote a culture of safety<br>• Address common system flaws that enable mistakes<br>—Lost tests<br>—Unavailable experts<br>—Communication barriers<br>—Weak coordination of care<br>• Provide cognitive aids and decision support resources<br>• Encourage consultation and second opinions<br>• Develop ways to allow effective and timely feedback |
| Patients | • Be good historians, accurate record keepers, and good storytellers<br>• Ask what to expect and how to report deviations<br>• Ensure receipt of results of all important tests |
| **Indirect and Supplemental** | Role |
| Oversight organizations | • Establish expectations for organizations to promote accurate and timely diagnosis<br>• Encourage organizations to promote and enhance<br>—Feedback<br>—Availability of expertise<br>—Fail-safe communication of test results |
| Medical media | • Ensure an adequate balance of articles and editorials directed at diagnostic error<br>• Promote a culture of safety and open discussion of errors and programs that aim to reduce error |
| Funding agencies | • Ensure research portfolio is balanced to include studies on understanding and reducing diagnostic error |
| Patient safety organizations | • Focus attention on diagnostic error<br>• Bring together stakeholders interested to reduce errors<br>• Ensure balanced attention to the issue in conferences and media releases |
| Lay media | • Desensationalize medical errors<br>• Promote an atmosphere that allows dialogue and understanding<br>• Help educate patients on how to avoid diagnostic error |

upcoming days, weeks, and months, and when and how to convey any discrepancies to the provider.

If there is no clear diagnosis, this too should be conveyed. Patients prefer a diagnosis that is delivered with confidence and certainty, but an honest disclosure of uncertainty and the probabilistic nature of diagnosis is probably a better approach in the long run. In this framework, patients would be more comfortable asking questions such as "What else could this be?" Exploring other options is a powerful way to counteract our innate tendencies to narrowly restrict the context of a case or jump too quickly on the first diagnosis that seems to fit.

### Be Watchdogs for System-related Errors

In a perfect world, all test results would be reliably communicated and reviewed, all care would be well coordinated, and all medical records would be available and accurate. Until then, the patient can play a valuable role in combating errors related to latent flaws in our healthcare systems and practices. Patients can and should function as back-ups in this regard. They should always be given their test results, progress notes, discharge summaries, and lists of their current medications. In the absence of reliable and comprehensive care coordination, there is no better person than the patient to make sure information flows appropriately between providers and sites of care.

## OTHER STAKEHOLDERS

Oversight organizations such as the Joint Commission recently have entered the quest to reduce diagnostic error by requiring healthcare organizations to have reliable means to communicate test results. Healthcare organizations by necessity pay attention to Joint Commission expectations; these expectations should be expanded to include the many other organizational factors that have an impact on diagnostic error, such as encouraging feedback pathways and ensuring the consistent availability of appropriate expertise.

Both the lay media and professional journals could further the cause of accurate and timely diagnosis by drawing attention to this issue and ensuring that diagnostic error receives a balanced representation as a patient safety issue. The media also must acknowledge a responsibility to promote a culture of safety by desensationalizing medical error. If there is anything to be learned from how aviation has improved the safety of air travel, it is the lesson of continuous learning, not only from disasters but also from simple observation of near misses. The media could substantially aid this effort in medicine by emphasizing the role of learning while deemphasizing the emphasis on blame.

Thus far, funding agencies have underemphasized diagnostic error in favor of the many other aspects of the patient safety problem. This type of error is not regarded as one of the low-hanging fruit.[7] Although diagnostic error is estimated to cause an appreciable fraction of the adverse events related to medical error,[1] funded grants related to diagnosis are scarce. An obvious problem is that the solutions are less apparent for diagnostic errors than other types of mistakes (e.g., improper medication), so perhaps this imbalance simply reflects a lack of grant applications. If the funding were available, applications would follow.

Patient safety organizations could play a substantial role in advancing diagnostic accuracy and timeliness simply by bringing attention to this issue. This could take the form of dedicated conferences, or perhaps simply advancing diagnostic error as a featured theme at patient safety conferences and gatherings. In addition to drawing attention to the problem, these forums play an invaluable role in bringing together people interested in solutions, thus allowing for networking and synergies that can more rapidly lead the field forward.

## CONCLUSION

In summary, the faint blip of diagnostic error is finally growing stronger on the patient safety radar screen. An increasing number of publications are drawing attention to this issue. Research studies are starting to appear that use human factors approaches, observational techniques, or health services research protocols to better understand these errors and how to address them. In the proper order of things, our knowledge of diagnostic error will increase enough to suggest solutions, and patient safety leaders and leading healthcare organizations will begin to outline goals to reduce error, measures to achieve them, and monitors to check progress. A measure of progress will be the extent to which both physicians and patients come to understand the key roles they each can play to reduce diagnostic error rates. For the good of all those who are affected by diagnostic errors, these processes must start now.

## Acknowledgements

This work was supported in part from a grant from the National Patient Safety Foundation. We are grateful to Eta Berner, EdD, for review of the manuscript and to Grace Garey and Mary Lou Glazer for their assistance.

Mark L. Graber, MD
*Veterans Affairs Medical Center, Northport, New York, and*
*Department of Medicine,*
*State University of New York at Stony Brook*
*Stony Brook, New York, USA*

## AUTHOR DISCLOSURES

**Mark L. Graber, MD,** has no financial arrangement or affiliation with a corporate organization or a manufacturer or provider of products discussed in this article.

## References

1. Berner E, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(suppl 5A):S2–S23.
2. Graber ML, Franklin N, Gordon RR. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165:1493–1499.
3. Schiff GD. Introduction: communicating critical test results. *Jt Comm J Qual Patient Saf.* 2005;31:63–65.
4. Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Doi K. Relative gains in diagnostic accuracy between computer-aided diagnosis and independent double reading. *SPIE Med Imaging.* 2000;3981:10–15.
5. Garg AX, Adhikari NKJ, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA.* 2005;293:1223–1238.
6. Case and commentaries. AHRQ [Agency for Healthcare Research and Quality] Web M&M, January 2008. Available at: http://www.webmm.ahrq.gov. Accessed January 30, 2008.
7. Graber ML. Diagnostic error in medicine: a case of neglect. *Jt Comm J Qual Patient Saf.* 2005;31:106–113.