



Overconfidence as a Cause of Diagnostic Error in Medicine

Eta S. Berner, EdD,^a and Mark L. Graber, MD^b

^aDepartment of Health Services Administration, School of Health Professions, University of Alabama at Birmingham, Birmingham, Alabama, USA; and ^bVA 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. © 2008 Elsevier Inc. All rights reserved.

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 Lowry¹

Mongerson² 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 colleagues³ 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.

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.

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.

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, Elstein¹⁶ 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)
Pertlis (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 colleagues³² 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 associates³² 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 preva-

lent 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 colleagues⁷⁷ 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 colleagues⁷⁹ 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 colleagues⁸⁰ 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 au-

topsy, 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 Gawande⁸³

“He who knows best knows how little he knows.”

—attributed to Thomas Jefferson⁸⁴

“Doctors think a lot of patients are cured who have simply quit in disgust.”

—attributed to Don Herold⁸⁵

As Kirch and Schafii⁷⁸ 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 colleagues⁹⁰ 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 non-compliance.^{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 colleagues¹⁰⁰ 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 associates¹⁰¹ implemented com-

puter-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 Mauck¹⁰² 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, Tanenbaum¹⁰³ worries that evidence-based medicine will decrease the “professionalism” of the physician. van der Sijs and colleagues¹⁰⁴ 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 Katz¹⁰⁵ 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 colleagues¹⁰⁸ 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 colleagues¹¹¹ 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 antemortem were wrong 40% of the time.¹¹¹ Similar findings were reported by Landefeld and coworkers¹¹²; 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. Potchen¹¹³ 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 Croskerry¹⁰⁷ calls these responses our “cognitive predispositions to respond.” These heuristics are powerful clinical tools that allow problems to be solved quickly and, typi-

cally, 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 Norman¹³⁰ and Klein¹²⁸ 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. Klein¹²⁸ refers to this process as “recognition primed” decision making, referring to the extensive experience of the expert with previous similar cases. Gigerenzer and Goldstein¹³⁶ 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.

Hamm¹³⁷ 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

characteristic 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. Gigerenzer¹³⁸ (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 Norman¹³⁰ 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 phenomenon 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, 1871¹⁴⁵

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. Klein¹²⁸ 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. Theoretically, 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 colleagues¹⁰⁸ 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 Bordage¹⁵² 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’s¹²⁸ “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 colleagues¹⁵⁴ advocate this strategy; their definition of types of situational awareness is similar to what others have called metacognitive skills. Croskerry^{115,155} and Hall¹⁵⁶ 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.

Croskerry¹⁵⁷ 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 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 Nolan¹⁶⁴ 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 colleagues¹³⁹ performed a systematic evaluation of 4 of these systems; in 1994, Miller¹⁶⁵ 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 editorial¹⁶⁶ that accompanied the article by Berner and associates¹³⁹ is illustrative of an overall negative attitude toward these systems. In a subsequent study, Berner and colleagues¹⁶⁷ 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 associates¹⁷² 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 Miller¹⁷⁹ and Berner¹⁸⁰ 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} Vir-

tually 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 colleagues¹⁸² 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 disease¹⁴⁶—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 β -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 colleagues¹⁹⁴ about the Pap smear program, and Hall¹⁹⁶ 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 Redelmeier⁷⁶ and Gandhi et al⁷⁷ 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 colleagues²⁰⁷ 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 associates²⁰⁸ 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 Gladwell²¹¹ 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

- Lowry F. Failure to perform autopsies means some MDs “walking in a fog of misplaced optimism.” *CMAJ*. 1995;153:811–814.
- Mongerson P. A patient’s perspective of medical informatics. *J Am Med Inform Assoc*. 1995;2:79–84.
- 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.
- YouGov survey of medical misdiagnosis. Isabel Healthcare—Clinical Decision Support System, 2005. Available at: <http://www.isabel-healthcare.com>. Accessed April 3, 2006.
- Burroughs TE, Waterman AD, Gallagher TH, et al. Patient concerns about medical errors in emergency departments. *Acad Emerg Med*. 2005;23:57–64.
- Tierney WM. Adverse outpatient drug events—a problem and an opportunity. *N Engl J Med*. 2003;348:1587–1589.
- Norman GR, Coblenz CL, Brooks LR, Babcock CJ. Expertise in visual diagnosis: a review of the literature. *Acad Med*. 1992;67(suppl):S78–S83.
- Foucar E, Foucar MK. Medical error. In: Foucar MK, ed. *Bone Marrow Pathology*, 2nd ed. Chicago: ASCP Press, 2001:76–82.
- Fitzgerald R. Error in radiology. *Clin Radiol*. 2001;56:938–946.
- Kronz JD, Westra WH, Epstein JI. Mandatory second opinion surgical pathology at a large referral hospital. *Cancer*. 1999;86:2426–2435.
- Berlin L, Hendrix RW. Perceptual errors and negligence. *Am J Radiol*. 1998;170:863–867.
- 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.
- Neale G, Woloschynowych J, Vincent C. Exploring the causes of adverse events in NHS hospital practice. *J R Soc Med*. 2001;94:322–330.
- 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.
- 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.
- Elstein AS. Clinical reasoning in medicine. In: Higgs JJM, ed. *Clinical Reasoning in the Health Professions*. Oxford, England: Butterworth-Heinemann Ltd, 1995:49–59.
- 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.
- McGinnis KS, Lessin SR, Elder DE. Pathology review of cases presenting to a multidisciplinary pigmented lesion clinic. *Arch Dermatol*. 2002;138:617–621.
- Zarbo RJ, Meier FA, Raab SS. Error detection in anatomic pathology. *Arch Pathol Lab Med*. 2005;129:1237–1245.
- 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.
- Harris M, Hartley AL, Blair V, et al. Sarcomas in north west England. I. Histopathological peer review. *Br J Cancer*. 1991;64:315–320.
- 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.
- Goddard P, Leslie A, Jones A, Wakeley C, Kabala J. Error in radiology. *Br J Radiol*. 2001;74:949–951.
- 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 Healthcare 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 unnecessary 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, Wulffing 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, Schechter 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 Healthcare 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 probabilistic 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 Medicine*, 4th ed. St Louis, MO: Mosby, 1998:223–244.
70. Failure to diagnose. Mideagnosis of conditions and diseases. 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 Sijts 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, Gabršček 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 Educ*. 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, Schmitz 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, Zininger 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, Kassarian 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.