Why Did I Miss the Diagnosis? Some Cognitive Explanations and Educational Implications

GEORGES BORDAGE

Why did you miss the diagnosis in the past year? Ten community-based internists attending a continuing education activity, each of whom reported an average of four diagnostic errors in the past year, responded (in decreasing order):

It never crossed my mind.
I paid too much attention to one finding, especially lab results.
I didn't listen enough to the patient's story.
I was too much in a hurry.
I didn't know enough about the disease.
I let the consultant convince me.
I didn't reassess the situation.
The patient had too many problems at once.
I was influenced by a similar case.
I failed to convince the patient to investigate further.
I was in denial of an upsetting diagnosis.

In retrospect, the types of diagnostic errors reported by these internists represent a good cross section of the diagnostic errors described in the literature. Some are cognitive in nature, others are personal, and still others are organizational.

The purpose of this paper is to describe and understand diagnostic errors. However, the field is too broad to tackle each and every type of error. Thus, after presenting an overview, I focus on two error types: faulty detection of clinical features and faulty triggering of diagnostic hypotheses, both related to the initial aspects of a clinical encounter. Finally, and based on the lessons learned, I present some practical implications for planning and organizing introduction to clinical medicine (ICM) courses that help students optimally gather and interpret data and thus avoid detection and triggering errors.

Over the past three decades, an average of 35 papers have been published each year on diagnostic errors, according to a Medline search from 1966 to 1998. The range of topics is broad: from legal issues to psychosocial considerations; from visual perception to cognitive biases. For the purpose of this paper, I have excluded legal topics and errors related to visual diagnosis (see Norman's 1992 review[1]) and focused mainly on errors of cognition, selectively reviewing a total of 188 papers and a dozen books. Finally, in an era of evidence-based health care, I was especially sensitive to supporting my assertions with evidence, albeit not in a meta-analytic way.

Relevance

Why should we pay attention to diagnostic errors? Diagnostic errors are prevalent and consequential and may not resolve with time alone. Three sources helped me get a sense of the prevalence and consequences of diagnostic errors; namely, autopsy reports, health services research, and surveys of physicians-in-training.

Since the 1950s, numerous studies in which autopsy results were compared over different eras have shown a constant and stable diagnostic error rate of about 10%.[2–4] Kirch and Schaffi[4] identified three leading categories of diagnostic errors revealed by autopsies done in a university hospital. They reported 10% misdiagnoses (clinical diagnosis not found at autopsy) with worsened patient prognosis; for example, the physician diagnosed a brain tumor but at autopsy it turned out to be an old cerebrovascular accident. Second, they found a 25% error rate for missed diagnoses (false negatives) with no prognostic relevance; that is, an unknown diagnosis disclosed at autopsy. Finally, a third category of false-positive diagnoses (10%) had no consequences on prognosis.

The four most frequent diagnoses implicated in errors were pulmonary embolism, myocardial infarction, cancers, and infections, in particular pneumonia. For some of these diagnoses, new diagnostic technologies such as ultrasonography or CT scanning have not reduced the rate of misdiagnoses. Occasionally (6% to 9%), "misinterpretation, technical errors, and overreliance on these new procedures . . . contributed directly to diagnostic errors." In one case, for example, a positive cholangiogram led to a radiologic diagnosis of cholangiosepsis; at autopsy it was discovered that the patient had a pulmonary embolus. According to Williamson,[5] in the late 1980s, up to half of the patients with pulmonary emboli were being missed; that is, about the same rate as in 1900.

The traditional history and physical (H&P) are, in many instances, winning the tug of war with those diagnostic technologies, providing more conclusive information than do the various technologies. In Kirch and Schaffi's[4] study of 400 autopsies, they showed that "the history and physical examination played an important role in the diagnostic process, leading to a correct final diagnosis in 60% to 70% of cases" compared with 35% for imaging techniques. Chimowitz's[6] study of bedside diagnosis in neurology showed similar results when comparing diagnostic conclusiveness of the H&P with that of laboratory tests: "the diagnostic burden still fell on the clinician in more than 50% of cases." They concluded: "interviewing and examination skills, clinical knowledge, and reasoning abilities . . . remain our most powerful diagnostic tools." There is no doubt that diagnostic technologies, such as chemistry analyses in thyroid diseases or scans in head trauma and tumor location, are clinically useful. Even scans in diagnosing appendicitis could be advantageous[7]; but even then, as the editor pointed out, "skills in history taking and physical examination should be increased in order to determine better which patients need such scanning."[8]

Data from health services research also point to important deficiencies in diagnosis. Brook,[9] in an analysis of the literature on acute medical care for the elderly, found that 26% to 64% of elderly patients were subject to one or more serious diagnostic errors. Williamson,[10] in a study of the quality of ambulatory care in rural areas, found that 50% to 58% of patients were subject to a diagnostic error. An entire issue of the *American Journal of Medical Quality* (Winter 1994) was devoted to reports from a symposium on quality issues in diagnosis, pointing to the need for information synthesis of specific diagnostic literature.

Finally, what about diagnostic errors during training? Wu[11] asked house officers in three internal medicine programs in large tertiary care centers to describe their most significant mistake in the last year, that is, a mistake "with serious or potentially serious consequences for the patient and {that} would have been judged wrong by knowledgeable peers at the time it occurred." There was no effect due to centers, and the most frequent mistake reported was related to diagnosis (38 of 144 incidents, 33%). Important consequences included death (45%), delayed treatment (16%), delayed diagnosis (10%), and complications (8%). Noteworthy in Wu's study is the fact that only half of the house officers (54%) discussed the mistakes with their attending physicians.

Types of Diagnostic Errors

What went wrong? Why? The classification of human errors, and diagnostic errors in particular, tends to be very pragmatic. Senders and Moray[12] suggested three main categories of human errors. The first category (phenomenologic taxonomies) describes quite directly the events as observed (e.g., omission). The second category refers to the cognitive mechanisms involved (e.g., perception, memory, attention), and the third, to the biases and deep-rooted tendencies underlying the errors (e.g., confirmation bias).

Based on the literature reviewed, I classified diagnostic errors into three functional categories: (1) data-gathering errors (from observation to findings), (2) data-integration errors (from findings to diagnoses), and (3) situational factors. The first two categories come from observer variability studies and are not meant to be mutually exclusive. Indeed, I show later that there is constant interaction between the two that can lead to diagnostic errors. The third category, that of situational factors, is more etiologic in nature, pointing to possible psychological, social, and organizational causes for the errors. Altogether I culled 29 different types of diagnostic errors from the literature (see Table 1).[13–25]

Table 1. Types of Diagnostic Errors According to Data Gathering, Data Integration, and Situational Factors

Data gathering (from observation to findings)	
1.	Incomplete history of present illness or history and physical
2.	Ineffective questioning (interviewing)
3.	Failure to gather useful information to verify diagnosis
4.	Faulty detection
5.	Excessive data gathering
6.	Failure to validate findings with patient
7.	Misidentification of symptoms or signs
8.	Faulty or improper physical examination techniques
9.	Failure to screen
10.	Overreliance on someone else's history and physical
11.	Poor etiquette leading to poor data quality
12.	Misled by the way the information presented itself
Data integration (from findings to diagnoses)	
13.	Failure to consider a finding(s)
14.	Over- or underestimating the usefulness or meaningfulness of a finding(s)
15.	Faulty context formulation
16.	Faulty estimate of prevalence
17.	Failure to periodically review the situation
18.	Overreliance on someone else's opinion
19.	Reporting findings not gathered
20.	Faulty causal model: ignorance or misconceptions
21.	Failure to ask for advice (consultation)
22.	Failure to act sooner
23.	"No-fault" error: atypical case, extremely rare, or rapidly evolving
Situational factors	
24.	Stress
25.	Fatigue, too many hours
26.	Excessive workload, not enough time
27.	Physician uncomfortable with own feelings toward the patient
28.	Physician's mood or personality
29.	Work environment: equipment, support, peer pressure, rewards and punishment

Faulty or Improper Physical Examination Techniques

Faulty or improper physical examination techniques (error 8 in Table 1) are illustrative of errors in data gathering. For almost two thirds of the patients they examined, residents and interns in a general medicine service made at least one serious physical examination error—such as misinterpreting or omitting a purpura, a splenomegaly, or focal neurologic signs—that, once corrected, "frequently led to major changes in differential diagnosis and therapy."[16] One fourth to one third of the errors were attributable to incorrect findings (errors of commission), while the remainder were omissions. The worst resident made errors in almost every fourth finding observed. In large-scale, multi-site, multi-discipline studies of cardiac and pulmonary auscultation skills, Mangione[22,23,26] also found that physicians-in-training had great difficulty identifying ten commonly encountered respiratory events (mean error rate of 61%)[26] and 12 cardiac events (median error rate of 78%).[23] In both skills, the housestaff improved little with years of training. Similarly, in a study by St. Clair,[21] the overall error rates of medicine housestaff for three valvular diseases were high; 48% for mitral regurgitation, 63% for mitral stenosis, and 46% for aortic regurgitation. Both Mangione[27] and St. Clair advocated a parsimonious, selective approach to physical diagnosis, focusing on key, discriminating findings. St. Clair and colleagues found that "accurate recognition of a few 'key' observations was associated with a correct diagnosis in two of the three diseases"; for example, identifying a loud first heart sound, an opening snap, and an apical diastolic murmur for mitral stenosis. They concluded, "teaching housestaff to elicit and interpret a few critical signs accurately may improve their physical diagnosis abilities."[21] I address this issue later.

Over- or Underestimation of Findings

In the 1970s, Wiener and Nathanson[13] proposed five major categories of errors in physical diagnosis: technique, omission, detection, interpretation, and recording. Interpretation errors are well exemplified in failures to consider a finding (error 13 in Table 1) and over- or underestimation of the usefulness or meaningfulness of a finding (error 14 in Table 1).

The leading errors reported by the ten internists surveyed were related to ignoring or under- or overestimating clinical findings. Faulty estimation includes assigning the wrong weight or importance to a finding, using non-discriminatory or normal findings to support a diagnosis, and overly relying on laboratory data. The leading error among third-year medical students in the Friedman study[24] was the use of non-discriminatory findings to support their diagnoses (85%), a phenomenon also observed by Gruppen[28] among housestaff. For example, many medical students in the Friedman study believed that if Motrin® helped, their patient must have had osteoarthritis. But Motrin is not a good discriminator between osteoarthritis and, for example, rheumatoid arthritis. In psychological terms, this phenomenon is known as "pseudodiagnosticity,"[29] that is, the use or selection of suboptimal, irrelevant information to make a decision. Gruppen and colleagues concluded that "physicians appear to have difficulty recognizing the diagnosticity of information, which often results in decisions that are pseudodiagnostic or based on diagnostically worthless information."[28] They added, "faulty diagnostic decisions may more often result from difficulties people have in knowing what is useful, diagnostically meaningful information, rather than from difficulties in combining, integrating, and organizing available information to make categorical decisions."

Voytovich[14,30] was among the first to propose a classification of faulty data integration, including wrong synthesis, premature closure, inadequate synthesis, and omission. Premature closure was highly prevalent (91%) and was independent of level of training.[31] A decade later, Kassirer and Kopelman[19] proposed a five-category classification of cognitive diagnostic errors; namely, faulty triggering, faulty context formulation, faulty information gathering and processing, faulty verification, and "no-fault" errors. The last category refers to atypical or nonspecific cases, extremely rare cases, or rapidly evolving diseases, where clinicians would be likely to miss the diagnosis. (A word of caution: there is little commonality among the existing classifications of diagnostic errors, yet certain terms refer to the same phenomenon, such as premature closure, faulty verification, and overinterpretation. In the future, one should avoid inventing new terms to describe what is already defined elsewhere under a given expression. Using existing terms with proper explanations is preferable and less confusing.)

Some recurring themes emerged from the literature on faulty data gathering and integration. Time alone does not lessen the incidence of certain errors; diagnostic errors can have important adverse consequences (e.g., delayed or inappropriate treatment, unnecessary investigations, complications, or patient anxiety); diagnosis rests on a few key discriminating findings; and house officers are reluctant to show their thinking. I now turn your attention to some of the cognitive mechanisms related to missing a diagnosis, either because the finding was not detected or because it was not properly interpreted, and how these two processes may be related.

Understanding Diagnostic Errors

Consider the following patient taken from one of Norman's studies.[32] Mr. Geoffrey, a young man, comes to the emergency room for double vision and difficulty swallowing. Clinician A examines Mr. Geoffrey but doesn't see the ptosis, implying a detection or perception problem. Clinician B sees the ptosis, but never considers myasthenia gravis, a disease with which he is nonetheless familiar. This implies a data-processing or interpretation problem. When I presented this situation to the ten internists mentioned at the beginning of this paper and asked them for educational advice, they all responded by saying that clinicians need to be more thorough. They assumed that if you look long enough, the answer jumps at you. However, 20 years ago, Elstein, Shulman, and Sprafka[33] showed that thoroughness is not a good predictor of diagnostic accuracy!

Faulty detection or triggering can be attributed to a number of causes: having no prior instances to refer to ("I've never seen this before"), overestimating certain features ("Focused on the double vision"), or having no mental representation of the problem overall ("Had no sense of what I was dealing with"). Hamm and Zubialde[34] used three metaphors that parallel the causes above to explain physicians' thinking. Physicians recognize patterns (prior instances), follow rules (analyzing features), and activate neural networks (exploring complex pathways and concepts). I explore each mechanism.

Pattern Recognition

In a series of experimental studies in dermatology, Norman and collaborators[35,36] showed that physicians essentially use two modes of thinking: pattern recognition and analytical thinking. The first mode—the clinician has seen it before—is fast (under ten seconds), automatic, and largely accurate (e.g., 66% for general practitioners and 87% for dermatologists[35]). The second mode, analytical thinking, is slower and more conscious. Errors in diagnosis were not due to inattention (e.g., not taking enough time to think about it), but failure to make connections and quickly recall similar prior instances stored in memory.

In estimating the probability of an event, such as a diagnosis, the mind can use an approximation strategy, known in psychology as the availability heuristic, whereby the probability of the event is related to the ease of remembering specific instances from memory. When mental availability is different from actual frequency, the availability heuristic can become a source of bias. Dawson and Arkes[37] illustrated the memorability of a salient, yet infrequent, event with the case of an older physician who, having himself been diagnosed as having chronic appendicitis, afterwards overestimated chronic appendicitis among his older patients who complained of nonspecific abdominal discomfort. (Dawson and Arkes present a useful review of cognitive biases overall.) Also, personal experiences are not always generalizable to a larger population. Detmer, Fryback, and Gassner[38] showed that surgeons with high mortality rates in their subspecialties (e.g., neurosurgeons) overestimated mortality rates for surgery in general, compared with surgeons with low rates (e.g., plastic surgeons). Norman[1] concluded about diagnostic errors that "there is a large perceptual component, and errors in diagnosis originate as much from this non-analytic, rapid, and large unconscious (perceptual) component as from the conscious search for individual features and the integration of features into a clinical decision."

Analytical Thinking and Discriminating Features

What if you can't find a pattern, are puzzled, or are faced with an ill-defined problem? Then you switch to analyzing the features of the case where availability of diagnoses is based on a fixed set of features and rules. Feature identification is partially dependent on category identification.[39] In Norman's study,[32] students and experienced physicians, once presented with Mr. Geoffrey's complaint, either were given an interpretation of the raw data (e.g., bilateral ptosis) and asked for a diagnosis, or were given the diagnosis (i.e., myasthenia gravis) and asked to identify features. In both cases, both the students and the experienced physicians increased their accuracy substantially (20% for diagnosis). Features were more evident when the diagnosis was also available. Feature identification is a strongly interactive process in which both features and diagnoses are "co-selected," where the "suggested diagnosis focuses attention on specific features."[40]

In other words, you see what you are looking for. When the students and physicians were in the right ballpark (considering the correct diagnosis), they saw 59% of the features. When they were looking for the ballpark (making a differential diagnosis), they still saw more (50%) than when they were simply scouring the sports pages (thoroughly gathering more data; 44%). And when they were in the wrong park altogether (considering the wrong diagnosis), they saw only 40% of the features. Thus, thoroughness for the sake of thoroughness as an initial data-gathering strategy is suboptimal. Hatala[41] also showed that gathering more data (in that case, EKG features) not only did not improve diagnostic accuracy but was actually more an indication of diagnostic uncertainty.

So, if you don't know what you are looking for, there is a good chance you won't see it no matter how thoroughly you look. When in doubt, step back and view the problem from a different angle; gather and interpret data with another diagnosis or differential diagnosis in mind. ICM courses should teach students to conduct physical examinations while keeping three elements in mind simultaneously; namely, the technique (e.g., the maneuvers to examine the shoulder), some prototypical diagnoses (e.g., bursitis, rotator cuff tear, and osteoarthritis), and some key features or findings (e.g., localized versus generalized tenderness of the shoulder).

Students are better at interpreting available findings than selecting useful ones.[21,24,25,28] Wigton and collaborators[42,43] used policy-capturing methods to identify the discriminating features in various clinical situations. For example, for strep throat, they showed that the diagnosis best rests on four key findings: swollen anterior cervical nodes, tonsillar exudate or red throat, absence of cough, and history of fever.[44] This suggests that ICM courses should move away from teaching countless head-to-toe maneuvers (as many as 150) to impart a much smaller, more focused, and discriminating set of symptoms and signs. This tighter, more parsimonious knowledge would provide a solid foundation upon which to build competing diagnoses. This echoes Mangione and Peitzman's recommendation "to separate the wheat from chaff and discard signs and maneuvers of little value to the medical practice of the 1990s."[27]

Wigton[45–47] also showed that students benefit differently depending on the type of feedback they receive about their data-selection strategies. Simple outcome feedback (i.e., telling them that they've arrived at the correct or incorrect diagnosis) failed to improve students' learning. Instead, the students improved, and transferred their knowledge in practice, when they received "cognitive feedback," that is, when the optimal, discriminating findings were highlighted along with the actual weights used by the students to judge the importance of each clinical finding. The effect was greatest when the students were also provided with probability outcomes whereby they could also calibrate themselves over a set of cases (i.e., estimating base rates and posterior probabilities). Highlighting the findings not only pointed to the optimal, most discriminating findings, but also pointed to redundant findings (e.g., exudate and red throat are equivalent, not additive) and irrelevant (pseudodiagnostic) findings. This is a fertile ground for computer-aided simulations, where many cases can be presented to students in a short period of time. It allows for practice and feedback, two essential ingredients for learning.

Problem Representation

What if you want to explore or find explanations? You search networks of concepts and examples stored in memory. One strategy is to begin by getting an overall sense of the problem. Brenner and collaborators[48] showed that students (in their case, pre-algebra students) tried to solve problems before they had built representations of the problems in their minds. They were too close to the details (the numbers) to see the "big" picture (the relevant concepts and equations). Problem representation can be mediated in a number of ways, such as abstractions (e.g., acute–chronic, recurrent–unique), pictures (e.g., a mental image of the knee joint and the surrounding structures), charts, diagrams, or equations (e.g., acid–base balance). Studies in psychology have shown that with experience, mental representation becomes progressively more abstract and more remote from the actual data.[49]

When problem representation is viewed as abstractions, it might look like this: "Here's an *older* man with *acute, recurrent nocturnal* attacks of *severe* knee pain in a *single, large* joint, a *mono* arthritis." Patients do not come in talking this way. Instead, physicians transform findings into more abstract terms to define the type of problem or build a representation of the problem overall (e.g., an *episodic-mono* arthritis as opposed to a *chronic-poly* arthritis). My collaborators and I[50–55] have studied the use of abstractions by students and experienced clinicians in various clinical domains (i.e., neurology, gastroenterology, intensive care, and rheumatology) using complex problems and thinking-aloud procedures. Each transformation from a finding to an abstraction is called a semantic qualifier; for example, "last night" becomes "*acute*" onset (as opposed to *chronic*) or "right knee becomes "*single, large*" joint (as opposed to *multiple small* joints). Embedded in each qualifier is an opposing quality (e.g., *acute-chronic, mono-poly*) that can be used to actively compare and contrast diagnoses (e.g., a *single* and *large* joint is more compatible with gout than with rheumatoid arthritis or osteoarthritis). Successful diagnosticians used more semantic qualifiers in their discourses (referred to as elaborated discourses) than do unsuccessful ones.

Chang[56] studied the use of semantic qualifiers as a means of representing the chief complaint in a case of gout. The chief complaint was divided into six traditional attributes[57] (onset, site, course, severity, setting, and patient characteristics) and corresponding semantic qualifiers were identified (e.g., *sudden, acute* onset; *episodic, recurrent* course; *large, mono* joint). The clinicians with the correct diagnosis used more than twice as many basic semantic qualifiers (abstractions) than did the ones with the incorrect diagnosis (3.8 versus 1.6, $p < .04$). The clinicians did not transform all six elements; only three or four of the six were apparently sufficient to get a sense of the problem overall.

In problem representation, clinicians can use the semantic qualifiers as a memory index to riffle through concepts and instances in search of relevant diagnoses (a notion akin to Riesback and Shanck's[58] discrimination networks). For example, an *acute-mono* arthritis, as opposed to a *chronic-poly* arthritis, points to infectious or crystal-like problems that, in turn, point to septic arthritis as opposed to gout or pseudo-gout. The semantic qualifiers can also represent links to the basic sciences. For example, if a student puts joint fluid under the microscope, the presence of polymorphonuclear leukocytes (an early inflammatory reaction) can be associated with an acute onset, compared to lymphocytes (a secondary reaction) with a chronic onset. Thus, two bodies of knowledge come together, the clinical and histopathologic, where one can be used to construct explanations for the other.

Implications for ICM

What can be learned from our diagnostic errors? First, in many instances the error does not resolve with time alone, thus we need to intervene early—during the ICM course, for example. Second, practice is an essential element in preventing and recovering from errors. It has been long established in educational psychology that practice is the best predictor of performance.[29,59] Johnson and Carpenter[60] showed that housestaff proficiency in physical examination skills was directly correlated with the amount of time attending physicians spend at the bedside reevaluating physical findings. Short, sporadic practice sessions are not likely to succeed.[61] And last, self-recognition, assuming responsibility for mistakes, and sharing emotions lead to more constructive changes in practice.[11] Students and housestaff are reluctant to show their thinking,[11,62,63] and faculty seldom observe them directly[64,65] and tend to shy away from exposing errors.[66] Building a nonjudgmental and supportive environment is essential to decrease stress and future mistakes.[67] For example, Meyer[68] introduced readings and discussions in an ICM course for first-year medical students to openly address diagnostic errors and the fallibility of physicians, rather than trying to minimize errors.

Catching diagnostic errors early means starting at the introductory phase of physical diagnosis, that is, during the ICM course in most institutions. The goal is to get away from rote memorization of countless head-to-toe maneuvers and move toward building meaningful repertoires of examplars (patterns), recognizing and seeking key discriminating findings, and building conceptual understanding and indexing of problems (problem representation).

Building an initial repertoire of examples is a selective process. In a study in 1987,[69] I showed that "less is better to begin with" when building an initial fund of knowledge; that is, students learn better, at least at first, when they focus on prototypical disorders than when they spread themselves thin across a wide spectrum of causes. There was an inverse relationship between the number of disorders presented in class and the extent of prototype formation in the students' mind. One way to build a solid foundation of diagnoses is to focus initially on a limited set of prototypical disorders (e.g., sprain, compression fracture, or herniated disc in the case of low back pain) along with one or two less prototypical, yet serious, conditions (e.g., pyelonephritis or metastases). With 23 chief complaints (about 80% of the chief complaints in our outpatient department) and three to four disorders per complaint, ICM students could comfortably sort out 80 to 90 diagnoses—a good foundation to build on during subsequent clerkships and specialty training.

Building patterns and repertoire go hand in hand with gaining conceptual understanding of the clinical problems as a whole. One way to foster representation is by asking students to summarize their cases in an abstract manner using one or two sentences, as the successful diagnosticians did in Chang's study.[56] For example, "Mr. Clark, a *young man*, 32, and *previously healthy*, presents with a *recent*, 24-hour onset of *immediate, sharp, right*-sided low back pain *on ex-*

ertion, radiating *below* the knee. This favors a herniated disc as opposed to a sprain." Four abstract qualifiers can go a long way in sorting out low back pain: An "*acute-immediate-below* the knee-*localized*" problem, as in a herniated disc, as opposed to an "*acute-delayed-above* the knee-*non systemic*" problem, as in an inflammatory reaction (see the H&P of low back pain[70] in *JAMA*'s series on "The Rational Clinical Examination"). Thus, the ICM student examines the patient's back, not simply in a mechanistic manner, but with diagnoses, discriminating features (Norman's co-selection), and a sense of the big picture in mind all at once. With the help of a cadre of real or simulated patients, students can actively store and retrieve knowledge, all the while rehearsing and building traces in memory for future ease of access and understanding.[71] The semantic links offer scaffolding to represent and access similarities and differences across the deep structure of problems, and thus should facilitate transfer of knowledge.[72] Furthermore, the web of multiple competing diagnoses allows students to consider alternatives that are built into their knowledge base (e.g., the opposing semantic qualifiers and the related competing diagnoses), an excellent antidote to the one-track-minded clinician (as in the availability bias[29]). Also, the mix of disorders offers a means of early calibration.[73] Computer simulations could be a useful adjunct to real or standardized patients by presenting numerous cases to students (variations on a theme), highlighting prototypical diagnoses, discriminating features, and providing cognitive feedback.[45] As new information is learned, it can be integrated meaningfully into relevant prior knowledge and thus ensure maximum retrieval.[72]

In conclusion, there is a parallel between the three main goals of this paper and the clinical process. The paper brings together errors; the clinician gathers symptoms and signs. The paper sought to understand these errors; the clinician explores the etiology and pathophysiology of diseases. The paper suggested ways to prevent and remedy the errors; the clinician treats and follows the patient. Remediation based on simple descriptions runs the risk of instituting symptomatic treatments (e.g., mechanical thoroughness in data gathering). Understanding the errors and targeting the remediation within an etiologic approach is more likely to address the root of the problem and lead to a more satisfying resolution (e.g., co-selection of diagnoses and discriminating features in data acquisition). Three challenges face medical educators and researchers today: to translate theory into practice, as in designing ICM courses to foster a solid foundation of prototypical diagnoses and discriminating features in the student's mind and to avoid thoroughness for the sake of thoroughness; to evaluate educational innovations rather than being complacent (e.g., how well does problem representation work?); and, to challenge current wisdom to create new knowledge. Thus a healthy cycle is created in medical education between theory and practice and vice versa.

The content of this paper was delivered as the 1998 Invited Address for the section on Research in Medical Education (RIME) of the Group on Educational Affairs (GEA) at the annual meeting of the Association of American Medical Colleges (AAMC) in New Orleans. The author is grateful to the members of the GEA-RIME committee for the invitation. The author dedicates this paper to the memory of Dr. Terry Mast, founding editor of *Teaching and Learning in Medicine*, who was a dear friend, an engaging and relentless educator, and an outstanding human being.

References

1. Norman GR, Coblentz CL, Brooks LR, Babcock CJ. Expertise in visual diagnosis: a review of the literature. Acad Med. 1992;67:S78–S83.
2. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three medical eras. N Engl J Med. 1983;308:1000–5.
3. Pelletier LLJ, Klutzow F, Lancaster H. The autopsy: its role in the evaluation of patient care. J Gen Intern Med. 1989;4:300–3.
4. Kirch W, Schaffi C. Misdiagnosis at a university hospital in 4 medical eras. Report on 400 cases. Medicine. 1996;75:29–40.
5. Williamson JW. Quality of current diagnostic performance: our most serious health care quality problem? [editorial] Am J Med Qual. 1994;9:145–8.
6. Chimowitz MI, Logigian EL, Caplan LR. The accuracy of bedside neurological diagnoses. Ann Neurol. 1990;28:78–85.
7. Rao PM, Rhea JT, Novelline RA, Mostafavi AA, McCabe CJ. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. N Engl J Med. 1998;338:141–6.
8. McColl I. More precision in diagnosing appendicitis. N Engl J Med. 1998;338:190–1.
9. Brook RH, Kamberg CJ, Mayer-Oakes A. Appropriateness of Acute Medical Care for the Elderly: An Analysis of the Literature. Santa Monica, CA: Rand Corp., 1989.
10. Williamson JW, Fehlauer CS, Gaiennie MA. Assessing quality of ambulatory care. Qual Assur Utilization Rev. 1991;6:8–15.
11. Wu AW, Folkman S, McPhee SJ, Lo B. Do house officers learn from their mistakes? JAMA. 1991;265:2089–94.
12. Senders JW, Moray NP. Human Error: Cause, Prediction, and Reduction. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.
13. Wiener S, Nathanson M. Physical examination. Frequently observed errors. JAMA. 1976;236:852–5.
14. Voytovich AE, Harper L, Rippey RM. Evaluating analytic reasoning in the management of the multiproblem patient: a system for attaining reliability and validity. Presented at the Research in Medical Education (RIME) Conference of the Association of American Medical Colleges, Washington, DC, 1979.
15. Bordage G, Allen T. The etiology of diagnostic errors: process or content? An exploratory study. Presented at the Research in Medical Education (RIME) Conference of the Association of American Medical Colleges, Washington, DC, 1982.
16. Wray NP, Friedland JA. Detection and correction of house staff error in physical diagnosis. JAMA. 1983;249:1035–7.
17. Ornstein S, Johnson A, Markert G, Lawrence A. Association between family medicine residents' personality and laboratory test-ordering for hypertensive patients. J Med Educ. 1987;62:603–5.
18. Weinberg NS. The relation of medical problem solving and therapeutic errors to disease categories. Q Rev Biol. 1989;266–72.
19. Kassirer JP, Kopelman RI. Cognitive errors in diagnosis: instantiation, classification, and consequences. Am J Med. 1989;86:433–41.
20. Beaumier A, Bordage G, Saucier D, Turgeon J. Nature of the clinical difficulties of junior residents under direct observation. Can Med Assoc J. 1992;146:489–97.
21. St. Clair EW, Oddone EZ, Waugh RA, Corey GR, Feussner JR. Assessing housestaff diagnostic skills using a cardiology patient simulator. Ann Intern Med. 1992;117:751–6.
22. Mangione S, Nieman LZ, Gracely E, Kaye D. The teaching and practice of cardiac auscultation during internal medicine and cardiology training: a nationwide survey. Ann Intern Med. 1993;119:47–54.
23. Mangione S, Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. JAMA. 1997;278:717–22.
24. Friedman MH, Connell KJ, Olthoff AJ, Sinacore J, Bordage G. Medical student errors in making a diagnosis. Acad Med. 1998;73Suppl:S19–S21.
25. Gruppen LD, Palchik NS, Wolf FM, Laing TJ, Oh MS, Davis WK. Medical student use of history and physical information in diagnostic reasoning. Arthr Care Res. 1993;6:64–70.
26. Mangione S, Nieman LZ. Respiratory auscultatory skills of internal medicine and family practice trainees. Presented at the Eighth Ottawa International Conference, Philadelphia, PA, 1998.
27. Mangione S, Pietzman SJ. Physical diagnosis in the 1990s: art or artifact? J Gen Intern Med. 1996;11:490–3.
28. Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. Med Decis Making. 1991;11:233–9.
29. Wolf FM, Gruppen LD, Billi JE. Use of the competing-hypotheses heuristic to reduce 'pseudodignosticity.' J Med Educ. 1988;63:548–54.
30. Voytovich AE. The structured medical record as a tool in the analysis of student performance. Presented at the Research in Medical Education (RIME) Conference of the Association of American Medical Colleges, Washington, DC, 1975.
31. Voytovich AE, Rippey RM, Suffredini A. Premature conclusions in diagnostic reasoning. J Med Educ. 1985;60:302–7.
32. Norman GR, Brooks LR, Regehr G, Marriott M, Shali V. Impact of feature interpretation on medical student diagnostic performance. Acad Med. 1996;71:S108–S109.
33. Elstein AS, Shulman L, Sprafka S. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, MA: Harvard University Press, 1978.
34. Hamm RM, Zubialde J. Physicians' expert cognition and the problem of cognitive biases. Prim Care. 1995;22:181–212.
35. Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. Arch Dermatol. 1989;125:1063–8.
36. Allen SW, Norman GR, Brooks LR. Experimental studies in learning dermatologic diagnosis: the impact of examples. Teach Learn Med. 1992;4:35–44.

37. Dawson NV, Arkes HR. Systematic errors in medical decision making: judgment limitations. J Gen Intern Med. 1987;2:183–7.
38. Detmer DE, Fryback DG, Gassner K. Heuristics and biases in medical decision making. J Med Educ. 1978;53:682–3.
39. Norman GR, Brooks LR, Coblentz CL, Babcook CJ. The correlation of feature identification and category judgments in diagnostic radiology. Mem Cogn. 1992; 20:344–55.
40. Norman GR, LeBlanc V, Brooks LR. On the difficulty of noticing obvious features in patient appearance. Psychological Science. In press, 1999.
41. Hatala RA, Norman GR, Cunnington JP, Brooks LR. The effect of ECG instructional materials on medical students' reasoning strategy (personal communication). Department of Medicine, McMaster University, Hamilton, Ontario, Canada, 1998.
42. Wigton RS. Use of linear models to analyze physicians' decisions. Med Decis Making. 1988;8:241–52.
43. Tape TG, Heckerling PS, Ornato JP, Wigton RS. Use of clinical judgement analysis to explain regional variations in physicians' accuracies in diagnosing pneumonia. Med Decis Making. 1991;11:189–97.
44. Wigton RS, Connor JL, Centor RM. Transportability of a decision rule for the diagnosis of streptococcal pharyngitis. Arch Intern Med. 1986;146:81–3.
45. Wigton RS, Patil KD, Hoellerich VL. The effect of feedback in learning clinical diagnosis. J Med Educ. 1986;61:816–22.
46. Tape TG, Kripal J, Wigton RS. Comparing methods of learning clinical prediction from case simulations. Med Decis Making. 1992;12:213–21.
47. Poses RM, Cebul RD, Wigton RS, Centor RM, Collins, M, Fleischli G. Controlled trial using computerized feedback to improve physicians' diagnostic judgments. Acad Med. 1992;67:345–7.
48. Brenner ME, Mayer RE, Moseley B, et al. Learning by understanding: the role of multiple representations in learning algebra. Am Educ Res J. 1997;34:663–89.
49. Girotto V, Legrenzi P. Mental representation and hypothetico-deductive reasoning: the case of the THOG problem. Psychol Res. 1989;51:129–35.
50. Lemieux M, Bordage G. Structuralism and medical education: comparative studies of the cognitive strategies of novice physicians. Semiotic Inquiry. 1986;6:143–79.
51. Lemieux M, Bordage G. Propositional versus structural semantic analyses of medical diagnostic thinking. Cogn Sci. 1992;16:185–204.
52. Lemieux M, Bordage G. Comparing the core and the peel of the same fruit. Cogn Sci. 1993;17:143–7.
53. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. Acad Med. 1991;65Suppl:S70–S2.
54. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. Acad Med. 1994;69:883–5.
55. Bordage G, Connell K, Chang R, Gecht M, Sinacore J. Assessing the semantic content of clinical case presentations: studies of reliability and validity. Acad Med. 1997;72Suppl:S37–S39.
56. Chang RW, Bordage G, Connell KJ. The importance of early problem representation during case presentations. Acad Med. 1998;73Suppl:S109–S111.
57. Bates B, Bickley LS, Hoekelman RA. Physical Examination and History Taking. Philadelphia, PA: J. B. Lippincott, 1995.
58. Riesbeck CK, Schank RC. Inside Case-based Reasoning. Hillsdale, NJ: Lawrence Erlbaum Associates, 1989.
59. Chase WG, Chi MTH. Cognitive skill: implications for spacial skill in large scale environments. In: Harvey J (ed.). Cognition, Social Behavior, and the Environment. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.
60. Johnson J, Carpenter JL. Medical house staff performance in physical examination. Arch Intern Med. 1986;146:937–41.
61. Oddone EZ, Waugh RA, Samsa G, Corey R, Feussner JR. Teaching cardiovascular examination skills: results from a randomized controlled trial. Am J Med. 1993; 95:389–96.
62. Connell KJ, Bordage G, Chang RW, Howard BA, Sinacore JM. Measuring the promotion of thinking during precepting encounters in outpatient settings. Acad Med. 1999;74:S10–S12.
63. Baker EA, Bordage G, Connell KJ, Sinacore JM. Can diagnostic semantic competence be assessed from the medical record? Acad Med. 1999;74:S13–S15.
64. Stillman PL, Regan MB, Swanson DB, et al. An assessment of the clinical skills of fourth-year students at four New England medical schools. Acad Med. 1990; 65:320–6.
65. Stillman P, Swanson D, Regan MB, et al. Assessment of clinical skills of residents utilizing standardized patients: a follow-up study and recommendations for application. Ann Intern Med. 1991;114:393–401.
66. Ende J, Pomerantz A, Erickson F. Preceptors' strategies for correcting residents in an ambulatory care medicine setting: a qualitative analysis. Acad Med. 1995;70: 224–9.
67. Christensen JF, Levinson W, Dunn PM. The heart of darkness: the impact of perceived mistakes on physicians. J Gen Intern Med. 1992;7:424–31.
68. Meyer BA. A student teaching module: physician errors. Fam Med. 1989;21:299–300.
69. Bordage G. The curriculum: overloaded and too general. Med Educ. 1987;21:183–7.
70. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain. JAMA. 1992;268:760–5.
71. Regehr G, Norman GR. Issues in cognitive psychology: implications for professional education. Acad Med. 1996;71:988–1001.
72. Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. Acad Med. 1998; 73Suppl:S1–S5.
73. Fasoli A, Lucchelli S, Fasoli R. The role of clinical "experience" in diagnostic performance. Med Decis Making. 1998;18:163–7.