Commentary: **How Can We Make Diagnosis Safer?**

Gordon D. Schiff, MD, and Lucian L. Leape, MD

Abstract

Diagnostic errors are common and are a leading cause of patient dissatisfaction and malpractice suits. Because of its traditional heavy reliance on memory and lack of standardization, the diagnostic process is particularly error prone. A study by Zwaan and colleagues on diagnostic failures in treating dyspneic patients makes several important contributions: examining the process behind the diagnosis, seeking insights as to the reasons for the process failures by interviewing the treating physicians, and using the Delphi process with experts to map the optimal diagnostic process.

There is considerable confusion about definitions in the field of diagnostic errors. The authors of this commentary use a Venn diagram to clarify distinctions and relationships between diagnosis processes errors, delayed diagnosis and misdiagnosis, and adverse outcomes. A key question is whether a much more rigorous process should be employed for diagnosis, specifically the routine use of algorithms or guidelines, and whether barriers to achieving it can be overcome.

The authors propose an alternate simpler approach: six-part checklists for the top 20 or 30 clinical symptoms or problems. The elements of these checklists for minimizing diagnostic errors include essential data elements, don't-miss diagnoses, red-flag symptoms, potential drug causes, required referral(s), and follow-up instructions.

These checklists could—and should—be developed by collaborative efforts of the main users, primary care physicians, and emergency physicians, working with specialist physicians on specific symptoms and diagnoses. Absent such professional commitment, progress in diagnostic accuracy is likely to be slow.

Because of difficulties in defining, identifying, and preventing them, diagnostic errors have been on the back burner of the patient safety agenda, but they are increasingly being recognized as important and overlooked safety concerns.[1,2] They are the leading type of medical error resulting in malpractice claims and are high on the list of patient-reported failures in health care.[3] From a safety and reliability perspective, the diagnostic process is predictably error prone. It relies heavily on human memory, lacks systematic feedback systems, is highly idiosyncratic with widespread practice variations, and is plagued with difficulties in sorting out the signal of rare serious diagnosis from the noise of common conditions.

Dyspnea is a perfect example of such a signal-to-noise challenge. It is a commonly encountered symptom with a variety of self-limited or psychological causes, yet its differential diagnosis includes a number of serious and life-threatening entities that are often difficult to distinguish. Therefore, the study by Zwaan and colleagues[4] examining patient records for diagnostic errors in the treatment of 274 Dutch patients with dyspnea is welcome and timely. It makes several important contributions. The first is to dig below the surface to try to understand ways the treating physicians' diagnostic processes fell short. The researchers wisely refrained from labeling these shortfalls as "errors," instead coining a new term—"suboptimal cognitive acts" (SCAs)—suggesting alternatives that represent more optimal diagnostic thinking and action. In their chart reviews, the authors found SCAs in two-thirds of records, averaging 3.5 per patient. Fortunately, most of these did not result in a serious diagnostic error or patient harm, but the frequency and variety of the defective processes is disturbing.

After the charts were reviewed and suboptimal diagnostic processes identified, the researchers went back to the clinicians both to share with them the potential errors as well as to solicit their insights into what happened and why. This second innovative step represents an important and much-needed contribution to patient safety. Better understanding of what happened and why from the perspective of the front-line participants is an essential but often-missing ingredient when studying medical errors. This study uses the insights gained from the clinician interviews to categorize the identified SCAs, so it does not spell out in detail what was learned in these interviews. Given the complexity of diagnostic errors in medicine, however, this step is essential for informing improvement efforts, particularly in understanding barriers to optimal diagnostic process and implementing effective countermeasures to overcome these barriers.

A final noteworthy contribution, one that formed the basis for their assessments of

Dr. Schiff is associate director, Center for Patient Safety Research and Practice, Division of General Internal Medicine, Brigham and Women's Hospital, Boston, Massachusetts, and associate professor of medicine, Harvard Medical School, Boston, Massachusetts.

Dr. Leape is adjunct professor of health policy, Department of Health Policy and Management, Harvard School of Public Health, Boston, Massachusetts.

Correspondence should be addressed to Dr. Leape, Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115; telephone: (617) 576-6533; e-mail: leape@hsph.harvard.edu.

the adequacy of the cognitive processes, was to convene an expert panel to explicitly map out the "optimal diagnostic process" for dyspnea using a Delphi method. Part old-fashioned algorithm, part long-forgotten quality criteria map, and part checklist, the authors devised a model for differentiating some of the key diagnoses to consider in a dyspneic patient. Unfortunately, despite its length and complexity, the algorithm falls short of definitive and comprehensive guidance. For example, consider the issue of differentiating congestive heart failure (CHF) from pulmonary disease. With this diagnostic dilemma, controversies abound in interpreting clinical findings and selecting and interpreting laboratory tests (e.g., brain natriuretic peptide) and imaging studies. Differentiating CHF from pulmonary diseases is hardly an academic question because treatment can be radically different. Thus, we need to be certain that any algorithm is sufficiently valid that we can reliably ask clinicians to follow it.

Figure 1 Venn diagram illustrating relationships between diagnostic process errors, delayed diagnoses/misdiagnoses, and adverse outcomes. Group A: Adverse outcome resulting from error-related misdiagnosis (e.g., pathology specimens erroneously mixed up [diagnostic process error], resulting in wrong patient being given diagnosis of cancer [misdiagnosis], who then undergoes surgery with adverse outcome [adverse event]). Group B: Delayed diagnosis or misdiagnosis due to process error (e.g., positive urine culture overlooked, thus urinary tract infection not diagnosed, but patient has no symptoms or adverse consequences). Group C: Adverse event due to misdiagnosis, but no identifiable process error (e.g., death from acute myocardial infarction, but no chest pain or other symptoms that were missed). Group D: Harm from error in diagnostic process (e.g., colon perforation from colonoscopy done on wrong patient), but no misdiagnosis.

Confusion About Definitions

There is considerable confusion about definitions and terminology in the field of diagnostic error. This was evident in a recent discussion with leaders in this field attempting to write a white paper clarifying definitions. What is the difference between misdiagnosis, missed or delayed diagnosis, and diagnostic errors, and what are the relationships between diagnostic process and outcomes? Many studies on diagnostic error fail to sort these out. To their credit, we believe Zwaan and colleagues correctly distinguish three key aspects of diagnostic error. We offer a Venn diagram to illustrate these distinctions (Figure 1) and help clarify the relationships between errors in diagnostic process, errors in diagnosis, and resulting adverse outcomes.[1]

In relation to Figure 1, Zwaan and colleagues' SCAs represent a subset of a larger pool of diagnostic process errors (large circle) which, for example, could also include switching or losing specimens in a lab, overlooking an abnormal test result, or failing to perform a critical part of the physical exam or elicit a significant element of the medical history. These familiar process failures may or may not lead to an error in the
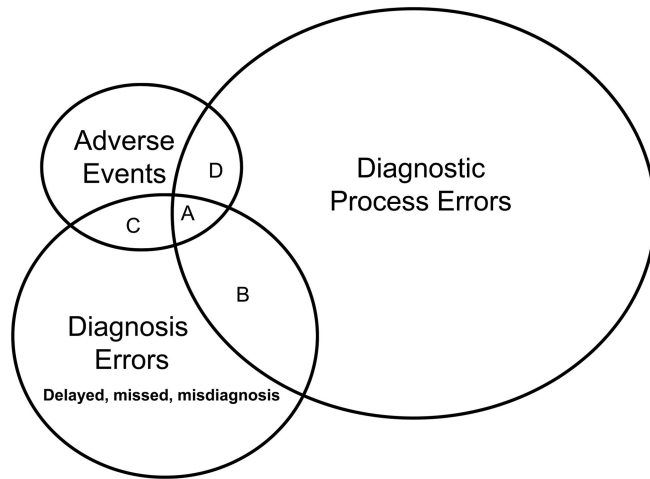
diagnosis (medium circle) and even less frequently lead to an adverse event (small circle, group A). The study by Zwaan and colleagues provides one of the first estimates of the frequency of such an intersection. Row 1 in Zwaan and colleagues' Table 3 corresponds to group "A" on the diagram. According to their analyses, nearly 3% of all records of patients presenting with dyspnea had evidence of an adverse event caused by a misdiagnosis resulting from a diagnostic process error.

Getting Real About Guidelines and Checklists

Should and could algorithms, such as the one Zwaan and colleagues used to guide their review, be prospectively deployed to reduce diagnostic errors by standardizing approaches to common clinical symptoms? This is hardly an original idea, but, given its obvious potential, it is worth asking why it has not yet happened. In this section, we examine some of the barriers to such efforts.

There is no perceived need

Most clinicians believe they are adequately guided by textbooks, review articles, and their training to know what needs to be done for a given patient. As

with other types of errors, most physicians have not measured and are not aware of their own diagnostic error rates. The data from Zwaan and colleagues confirm those from others that, lacking systematic feedback, many clinicians are inappropriately overconfident of the accuracy and appropriateness of their diagnostic actions and outcomes.[5]

It is too difficult to develop consensus

Developing consensus on which to base definitive recommendations is a complex and difficult undertaking, given controversies, complexities, and insufficient evidence. This may be true, but the experiences of Zwaan et al and many other researchers suggest it is time to try the experiment. Perhaps, rather than inflexible and all-encompassing algorithms, simpler, more manageable approaches (such as those we discuss below) would be fruitful and attainable.

A one-size-fits-all approach is not appropriate or desirable

Given the heterogeneous nature of patient characteristics, presentations, and even practice settings, there is no single algorithm that can apply to every situation. This is the standard, almost reflexive,

criticism of all attempts at standardization. But the purpose of standards, guidelines, and algorithms is not to force each patient into a rigid mold. Rather, it is to ensure that the physician does not overlook something essential to the evaluation. The point of an algorithm is not to prescribe treatment but, instead, to ensure that the clinician doesn't fail to consider appropriate treatment. Prospectively specifying standardized approaches provides clarity of what is supposed to happen, helping the clinician to spot ways a particular patient and his or her care differ from the norm.

There is no organization that is optimally positioned to produce and update such guidelines

There is some truth to this objection.[6] Specialty societies, one logical choice for developing diagnostic guidelines, often take a disease-specific approach or one narrowly oriented to a particular technology or test. Thus, guidelines from specialty societies exist for the diagnosis of asthma, for example, or the role of CT scanning in pulmonary embolism, or even evaluation of dyspnea in the cancer patient, but none exist on the general approach to evaluating a patient with dyspnea. Private health insurers issue guidelines, but their motives and credibility are often suspect.

One obvious national convener, the Agency for Healthcare Research and Quality, in an earlier incarnation as the Agency for Health Care Policy and Research, did undertake development of guidelines in the 1990s but was forced to end the project when Congress yielded to political pressure from specialty societies. A nongovernmental organization that might take responsibility for a renewed effort is the National Quality Forum. Representing multiple stakeholders, it has produced credible, widely supported, evidence-based, safe practices that are being used by hospitals nationwide. Diagnostic algorithms could be added to its charge.

Incorporating guidelines into clinical work flow is challenging

Even perfect guidelines are useless if they sit on the shelf (or Web site) and can't be used because they are too complex, time intensive, extensive, or expensive. The promise of automatically integrating guidelines into clinical decision support (CDS) programs is attractive but remains

List 1

Diagnosis Essentials: Elements of a Checklist for Minimizing Diagnostic Errors in Medicine

Essential data elements
Specific elements of history, physical examination, and testing data that should be reliably obtained for *every* patient presenting with a given symptom. In many situations, patients can reliably provide history elements with a computer-assisted questionnaire.

Don't-miss diagnoses
Critical diagnoses that can present with particular symptoms that are fatal or have serious consequences if they are not recognized and treated promptly. These diagnostic possibilities should be considered in every patient with such a symptom.

Red-flag symptoms
Specific associated symptoms or findings (e.g., back pain with new urinary incontinence in a cancer patient) that may indicate a serious condition and should lead to heightened suspicion of and evaluation for a don't-miss diagnosis.

Potential drug causes
Medications that can cause that symptom (or a disease manifesting that symptom). A high fraction of patient symptoms are medication side effects, yet medications are frequently not considered as a possible cause for symptoms. For this reason, medications need to be put at the top of the differential diagnosis for any unexplained symptoms. The computer should be able to assist in matching patients' medication profiles with known adverse effects.

Required referral(s)
When specialist expertise or technology is needed to adequately and safely evaluate the patient. Such cases may include possible rare conditions with which only specialists have sufficient experience, or conditions for which required testing (such as a biopsy or endoscopy) necessitates such a referral.

Patient follow-up instructions and plan
Warnings that patients should receive regarding specific symptoms that should lead them to return or call. These should be in writing and include a time frame (e.g., Call if you develop rash or fever or if you are not improved in 48 hours).

unfulfilled. The idea has floundered in a number of ways and for various reasons, including inadequacies of existing electronic health records (EHRs) (capabilities, work-flow design, data interoperability), poorly designed alerts (excess of false positives), failure to formulate guideline recommendation statements in "actionable" language (using "weasel" words instead of clear, executable statements), and an overall lack of attention to human factors and work flow. These limitations can and should be addressed in the next generation of guidelines, EHRs, and CDS programs.

How Can We Make Diagnosis More Reliable?

We believe that valid and usable comprehensive diagnostic algorithms and guidelines could be developed if adequate resources were provided, but it seems unlikely that this will happen in the near future. A more practical alternative would be simple rules developed in a joint effort by specialty societies and the main end users: primary care and emergency physicians. These simple rules could be developed as checklists for each of the top 20 to 30 critical symptoms. List

1 suggests six categories of information that could be included.

If professional societies developed such checklists in a standardized, consensus-based, and evidence-based fashion, they would be widely accepted and could be hard-wired into our care systems. As medical records are automated, adherence to the checklists can be easily tracked to produce outcomes-based evidence of their effectiveness and permit continuous improvement of the checklists themselves. Deploying this guidance widely—including putting this information into patients' hands—would create a type of safety net our health care system is currently lacking. Weaving these six elements into work flow, education, clinical culture, and electronic supports could make diagnosis much more reliable.

Enhancing our diagnostic infrastructure in such a thoughtful and collaborative way while at the same time avoiding overtesting, creating unwarranted patient fears, or increasing liability concerns will be challenging. But, absent such a serious professional commitment, we shouldn't hold our breath awaiting better diagnosis for dyspneic or any other patients.

References

1 Schiff GD, Kim S, Abrams R, et al. Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project. In: Henriksen K, Battles JB, Marks ES, Lewin DI, eds. Advances in Patient Safety: From Research to Implementation. Vol 2. Rockville, Md: Agency for Healthcare Research and Quality; 2005.

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

3 Bishop TF, Ryan AM, Casalino LP. Paid malpractice claims for adverse events in inpatient and outpatient settings. JAMA. 2011; 305:2427–2431.

4 Zwaan L, Thijs A, Wagner C, van der Wal G, Timmermans DRM. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. Acad Med. 2012;87:149–156.

5 Schiff GD. Minimizing diagnostic error: The importance of follow-up and feedback. Am J Med. 2008;121(5 suppl):S38–S42.

6 Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, eds. Clinical Practice Guidelines We Can Trust. Washington, DC: Institute of Medicine; 2011.