

Diagnosing Diagnostic Mistakes

May 1, 2005

McNutt RA, Abrams RI, Hasler S. Diagnosing Diagnostic Mistakes. PSNet [internet]. 2005.

<https://psnet.ahrq.gov/web-mm/diagnosing-diagnostic-mistakes>

Learning Objectives

- Understand the biases that may contribute to overcalling medical errors
- Describe the impact of considering the clinical spectrum of disease presentations or alternative diagnoses on assessment of error
- Appreciate the challenges inherent in assigning the label of “missed diagnosis” to a clinical scenario

The Commentary

In medicine, there is often confusion about which harms are reducible and which inevitable. Judgments about appropriateness are often based on incomplete, indirect, or unreliable data. At present, the knowledge base in safety research cannot provide definitive correlations between clinical decisions, the systems of delivery, and purported adverse events.

Given these deficiencies, we believe that—both in the literature and in practice—“error” is being overcalled. Some classifications consider the terms “missed diagnosis” or “delayed diagnosis” as merely adverse events (and not errors), reserving the designation of “error” for those missed or delayed diagnoses that are preventable. We think that the designation of “missed or delayed diagnoses” is fraught with ambiguity and unclear definitions. Applying these terms requires the use of judgment, and we believe that these judgments are often flawed, or at least debatable.

Error identification schemes often judge error as emanating from simple chains of events that can be understood and easily reconfigured. But, uncovering cause and effect is more complex than identifying simple relationships—redundancy and codependency abound.^(1,2) The dependencies among risk factors in safety research (both decisions and systems) are particularly likely to require rigorous scientific methods to determine cause and effect, because dependencies are so prevalent in complex systems.

Identifying diagnostic mistakes is especially difficult. As part of a grant from the Agency for Healthcare Research and Quality (AHRQ), our group has been working on evaluating diagnostic decisions to determine how many truly represent “errors.” As part of this research, we evaluated cases from the

archives of AHRQ WebM&M to determine whether we come to the same conclusion as the case discussants (and presumably the editors): namely, that a submitted case represents a preventable medical mistake. We have found that we often disagree with the attribution of patient harm to errors. In this commentary, we will review two cases from the WebM&M archives to illustrate some potential sources of ambiguity that influence the attribution of error.

Issue #1

Overcalling error due to evaluation of a case with knowledge of the patient's outcome (hindsight bias); especially in the face of no gold standard for diagnosis.

[Illustrative Case—Doctor, Don't Treat Thyself](#)

A 50-year-old radiologist presented with shortness of breath and interpreted his own chest x-ray as being “consistent” with pneumonia. Later the patient died of a myocardial infarction and pulmonary edema. Several radiologists reviewed the chest x-ray (after the outcome) and reported it “consistent” with pulmonary edema. The case was deemed by the discussant to “dramatically and tragically” illustrate a diagnostic mistake.

We don't believe that this case represents a preventable mistake or even a missed diagnosis, since, in our view, useful classification systems for preventable mistakes cannot include cases with no gold standard for the diagnosis. We recognize that this stringent standard, if widely applied, may suggest that diagnostic error cannot be defined by outcomes of care.

By “gold standard” we mean a high level of agreement about the criteria for diagnosis, such that different observers would agree when applying the criteria. The lack of agreement among observers trying to discriminate pneumonia from heart failure on a chest x-ray of a dyspneic patient is partially due to ambiguity in the criteria for making these diagnoses. For example, the overall accuracy of the evaluation of dyspnea is imprecise.⁽³⁾ Studies suggest that variation between observers when assessing patients accounts for at least some of the inaccuracy in assessing the value of classic signs and symptoms of heart failure.⁽⁴⁾ In one study, two university radiologists could identify chest x-ray signs of pneumonia with only “fair to good” reliability and were especially poor in defining the pattern of infiltrates.⁽⁵⁾

Given the ambiguity in the diagnostic criteria for heart failure and pneumonia and the imperfect agreement for interpreting the chest x-ray, radiologists who know a patient's outcome interpret chest x-ray findings differently than radiologists who are unaware of the ultimate outcome (hindsight bias).^(6,7) In the case at hand, it is irrelevant that the first radiologist was the patient; this discrepancy could occur between any two radiologists. Subjective criteria will be influenced by who, when, where, and how judgments are made.

The best way to minimize the impact of hindsight bias is to ignore the outcome. The outcome (good or bad) for a patient should not be revealed when trying to classify diagnostic mistakes. Evaluation of an adverse event should be done by an independent review panel following a structured, evidence-based assessment of the processes—not the outcomes—of care.⁽⁸⁾

Issue #2

Overcalling error due to failure to consider the spectrum of clinical presentations and the consequences of competing diagnoses.

Illustrative Case—Crushing Chest Pain: A Missed Opportunity

A 62-year-old woman is admitted with crushing chest pain and treated for possible myocardial infarction (MI). She later dies of an aortic dissection (AD). The case, when discussed on AHRQ WebM&M, was felt by the discussant to be a diagnostic error.

This case does not represent a missed diagnosis or a preventable error for several reasons. First, clinical presentations of a particular disease vary. Some dissecting aneurysms can be noted by the most casual of observations (i.e., wide and expanding mediastinum in a patient known to have an aneurysm), while some can be missed even after utmost scrutiny. In our judgment, this case is an example of the latter. The discussant claimed that the diagnosis was missed due in part to failure to note a small calcium deposit on the chest x-ray (again noted in hindsight), but the AHRQ WebM&M presentation had to magnify the finding on the radiograph in order to render it visible to readers ([Figure](#)). It is not clear how many observers, blinded to the outcome, would miss this finding, but surely some would. Before this case could be classified as an error, we would need to show that trained observers, blinded to the outcome, would see the radiologic finding and would, correctly and earlier in the patient's course, have made the diagnosis of dissecting aneurysm.⁽⁹⁾ Second, the quality of the literature addressing various diagnostic tests for AD is poor. Reports of the accuracy of diagnostic tests are biased by retrospective reviews of charts of patients known to have suffered from AD.⁽¹⁰⁻¹²⁾ Certainly those reports do not include patients requiring a magnified view of a small abnormality.

In addition to the failure to consider the spectrum of disease presentations, another reason for mislabeling is the failure to consider the consequences of competing diseases as the cause of a patient's complaints. The diagnostic process is not static; instead, it is a fluid refinement of possible diagnoses.⁽¹³⁾ Yet potential cases of missed or delayed diagnoses often come to light when a "surprise" diagnosis is found—one different from the diagnosis being pursued and/or treated. This surprise diagnosis is too often assumed to be "missed." This case illustrates that missing one diagnosis may be preferred to missing another.

The dissecting aneurysm case is one in which (i) several serious diseases may explain the patient's complaint; *and* (ii) empiric treatment of one may increase the chance of death in another; *and* (iii) the value of diagnostic tests to differentiate one disease from another is unknown or poorly studied. For example, MI, acute coronary syndrome (ACS), pulmonary embolus (PE), and AD are all in the differential diagnosis of the patient's complaints. These diseases take varying times to diagnose and have markedly different probabilities of occurrence, and each has different risks and benefits for diagnostic testing or delaying effective therapies.

The result of these uncertainties may require a physician to act on one disease to the detriment of another. For example, a workup for AD may delay life-saving anticoagulant therapy for someone who is having ACS. Since ACS is more common than AD in most clinical situations, more harm than good may come

from an overzealous attempt to not miss AD.

While this tradeoff has not been formally analyzed, we believe that such an analysis would conclude that AD represents a virtually insoluble diagnostic problem for many of its clinical presentations. To illustrate why, we performed a “back of the envelope” decision analysis. In it, we assumed that the ratio of AD:ACS for patients presenting with chest pain is at most 1:250.[\(14-16\)](#) We further assumed that if we found AD without delay (through the use of a perfectly accurate test), the patient’s life would be saved. We then assumed that a delay in anticoagulant therapy for ACS that occurs while assessing all patients for AD would lead to a 1% increase in death or myocardial infarction.[\(17,18\)](#) In summary, we assumed a 100% marginal improvement for making the diagnosis of AD and a 1% marginal improvement for early therapy for patients with ACS. These assumptions bias the analysis for making the diagnosis of AD (since, in real life, some people do die even when the diagnosis is made promptly). Given these assumptions, a vigorous search for AD in all patients that delayed anticoagulation for possible ACS would kill or cause MI in 2-3 patients with ACS for every patient with AD whose life would be saved. We use this threshold model of decision making as our criterion standard in diagnosis care evaluations.[\(19\)](#) This standard requires that explicit, evidence-based tradeoffs for competing diagnoses be considered before asserting that a preventable diagnostic mistake may have occurred.

Summary

Hindsight bias, especially in the face of subjective criteria for definitions of disease and test interpretation, a failure to consider the ambiguity imposed by the spectrum of clinical presentations for a single disease, and the complex trade-offs between competing causes of a patient’s complaints, makes classifying errors in diagnosis precarious. We believe that any attempt to determine cause and effect (preventability) must incorporate these factors [\(Table\)](#). We believe that diagnostic error may be a misused concept when outcomes of care are considered, and that clinical practitioners rarely fully appreciate their own limitations in analyzing their diagnostic reasoning, especially once the case’s outcomes are revealed.[\(20\)](#) Instead, error classification may need to shift to processes of care, asking such question as: Are the diagnosticians seeking a reasonable differential diagnosis? Do diagnostic plans incorporate the risk/benefit of finding one diagnosis rather than another? Were the appropriate tests ordered for the differential diagnosis list?

Research on diagnostic errors may be stifled until we have an accurate method to assign cause-and-effect relationships to our decisions, systems of care, and adverse events, supported by a robust taxonomy of diagnostic improvement opportunities. When cases are deemed to be diagnostic errors in journal articles (including AHRQ WebM&M) and elsewhere, it is worth considering contrary opinions, focused particularly on whether one or more of the four hypothesized classification conundrums has occurred [\(Table\)](#). Such an interchange may hasten the development of a useful classification scheme for diagnostic mistakes.

Robert McNutt, MD Professor and Associate Chair, Department of Medicine Associate Director, Medical Informatics and Patient Safety Research Rush University Medical Center, Chicago, IL

Richard Abrams, MD Associate Professor and Program Director, Department of Medicine Co-Chair, Patient Safety Committee Rush University Medical Center, Chicago, IL

Scott Hasler, MD Assistant Professor and Associate Program Director, Department of Medicine Rush University Medical Center, Chicago, IL

Faculty Disclosure: *Drs. McNutt, Abrams, and Hasler have declared that neither they, nor any immediate member of their families, has a financial arrangement or other relationship with the manufacturers of any commercial products discussed in this continuing medical education activity. In addition, they do not intend to include information or discuss investigational or off-label use of pharmaceutical products or medical devices.*

References

1. McNutt RA, Abrams R, Aron DC, for the Patient Safety Committee. Patient safety efforts should focus on medical errors. JAMA. 2002;287:1997-2001. [[go to PubMed](#)]
2. McNutt RA, Abrams RI. A model of medical error based on a model of disease: interactions between adverse events, failures, and their errors. Qual Manag Health Care. 2002;10:23-28. [[go to PubMed](#)]
3. Mulrow CD, Lucey CR, Farnett LE. Discriminating causes of dyspnea through clinical examination. J Gen Intern Med. 1993;8:383-392. [[go to PubMed](#)]
4. Badgett RG, Lucey CR, Mulrow CD. Can the clinical examination diagnose left-sided heart failure in adults? JAMA. 1997;277:1712-1719. [[go to PubMed](#)]
5. Albaum MN, Hill LC, Murphy M, et al. Interobserver reliability of the chest radiograph in community-acquired pneumonia. PORT Investigators. Chest. 1996;110:343-350. [[go to PubMed](#)]
6. Carthey J. The role of structured observational research in health care. Qual Saf Health Care. 2003;12(suppl 2):ii13-16. [[go to PubMed](#)]
7. Lilford RJ, Mohammed MA, Braunholz D, Hofer TP. The measurement of active errors: methodological issues. Qual Saf Health Care. 2003;12(suppl 2):ii8-12. [[go to PubMed](#)]
8. McNutt RA, Abrams R, Hasler S, et al. Determining medical error. Three case reports. Eff Clin Pract. 2002;5:23-28. [[go to PubMed](#)]
9. Berlin L. Defending the "missed" radiographic diagnosis. AJR Am J Roentgenol. 2001;176:317-322. [[go to PubMed](#)]
10. Klompas M. Does this patient have an acute thoracic aortic dissection? JAMA. 2002;287:2262-2272. [[go to PubMed](#)]
11. Hagan PG, Nienaber CA, Isselbacher EM, et al. The International Registry of Acute Aortic Dissection (IRAD): new insights into an old disease. JAMA. 2000;283:897-903. [[go to PubMed](#)]
12. Moore AG, Eagle KA, Bruckman D, et al. Choice of computed tomography, transesophageal echocardiography, magnetic resonance imaging, and aortography in acute aortic dissection: International Registry of Acute Aortic Dissection (IRAD). Am J Cardiol. 2002;89:1235-1238. [[go to PubMed](#)]

13. Kassirer JP, Kopelman RI. Learning Clinical Reasoning. Baltimore, MD: Lippincott Williams & Wilkins; 1991.
14. Meszaros I, Morocz J, Szlavi J, et al. Epidemiology and clinicopathology of aortic dissection. Chest. 2000;117:1271-1278. [[go to PubMed](#)]
15. Conti A, Paladini B, Toccafondi S, et al. Effectiveness of a multidisciplinary chest pain unit for the assessment of coronary syndromes and risk stratification in the Florence area. Am Heart J. 2002;144:630-635. [[go to PubMed](#)]
16. Khan IA, Nair CK. Clinical, diagnostic, and management perspectives of aortic dissection. Chest. 2002;122:311-328. [[go to PubMed](#)]
17. Husted SE, Kraemmer Nielsen H, Krusell LR, Faergeman O. Acetylsalicylic acid 100 mg and 1000 mg daily in acute myocardial infarction suspects: a placebo-controlled trial. J Intern Med. 1989;226:303-310. [[go to PubMed](#)]
18. Madsen JK, Pedersen F, Amtoft A, et al. Reduction of mortality in acute myocardial infarction with streptokinase and aspirin therapy. Results of ISIS-2. Ugeskr Laeger. 1989;151:2565-2569. [[go to PubMed](#)]
19. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109-1117. [[go to PubMed](#)]
20. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. N Engl J Med. 1989;320:1489-1491. [[go to PubMed](#)]

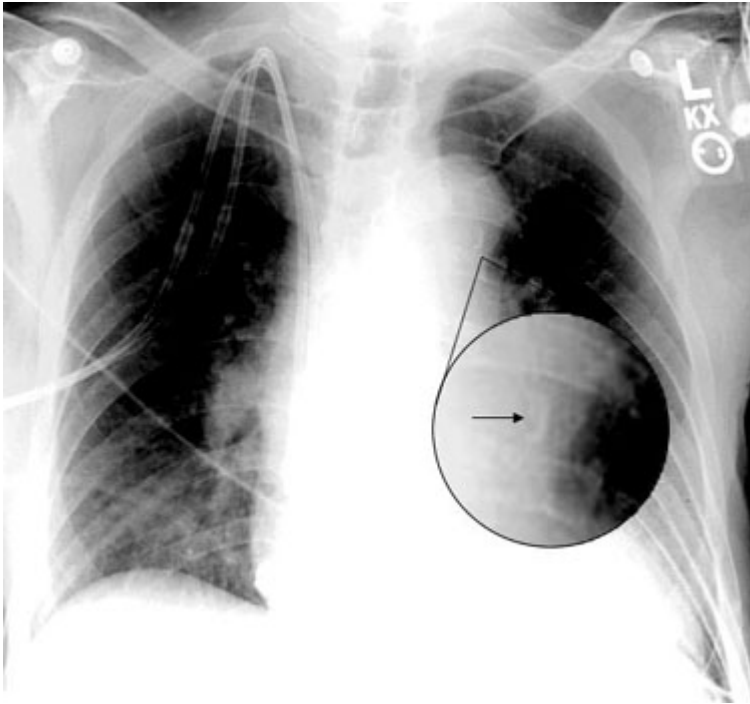
Table

Table. Hypothesized Classification Conundrums

- Hindsight bias—evaluation of a case with knowledge of the outcome
- Lack of a gold standard for the diagnosis
- Failure to consider the spectrum of clinical presentations
- Failure to consider competing diagnoses

Figure

Figure. Calcium Sign Magnified in the Radiograph Presented in Crushing Chest Pain Case



This project was funded under contract number 75Q80119C00004 from the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services. The authors are solely responsible for this report's contents, findings, and conclusions, which do not necessarily represent the views of AHRQ. Readers should not interpret any statement in this report as an official position of AHRQ or of the U.S. Department of Health and Human Services. None of the authors has any affiliation or financial involvement that conflicts with the material presented in this report. [View AHRQ Disclaimers](#)