The incidence of diagnostic error in medicine

Mark L Graber

ABSTRACT
A wide variety of research studies suggest that breakdowns in the diagnostic process result in a staggering toll of harm and patient deaths. These include autopsy studies, case reviews, surveys of patient and physicians, voluntary reporting systems, using standardised patients, second reviews, diagnostic testing audits and closed claims reviews. Although these different approaches provide important information and unique insights regarding diagnostic errors, each has limitations and none is well suited to establishing the incidence of diagnostic error in actual practice, or the aggregate rate of error and harm. We argue that being able to measure the incidence of diagnostic error is essential to enable research studies on diagnostic error, and to initiate quality improvement projects aimed at reducing the risk of error and harm. Three approaches appear most promising in this regard: (1) using ‘trigger tools’ to identify from electronic health records cases at high risk for diagnostic error; (2) using standardised patients (secret shoppers) to study the rate of error in practice; (3) encouraging both patients and physicians to voluntarily report errors they encounter, and facilitating this process.

In God we trust, all others bring data

The patient safety movement in the USA has entered its second decade. A wide range of important safety concerns have been studied, and to this point, including medication errors, hospital-acquired infections, wrong-site surgery and a host of other issues. Strangely lacking, however, is a concerted effort to find, understand and address diagnostic errors. Arthur Elstein, a cognitive psychologist interested in ‘how doctors think’, studied clinical decision making for his entire career and concluded the diagnosis is wrong 10–15% of the time. The incidence of diagnostic error has been estimated using eight different research approaches (table 1). Autopsy studies identify major diagnostic discrepancies in 10–20% of cases. Most cases in autopsy series derive from inpatient settings, but they also include deaths from the emergency department which, for many reasons, is considered to be the natural laboratory for studying diagnostic error. Although autopsies have virtually disappeared in the USA, autopsies are still common in many other countries, and despite the availability of modern imaging, continue to show diagnoses being missed that might have been lifesaving, particularly infections and cardiovascular conditions. Although autopsy data is considered the ‘gold standard’ in terms of providing the most definitive data on the accuracy of diagnosis, only a subset of cases ever reach autopsy, and in many cases, the relationship between clinical diagnoses and autopsy findings remains unclear.
Autopsies also discover a large number of incidental findings that were not suspected during life, but that were clinically irrelevant.

Surveys have found that diagnostic errors are a major concern of both patients and physicians. A survey of over 2000 patients found that 55% listed a diagnostic error as their chief concern when seeing a physician in an outpatient setting. Similarly, physician surveys have consistently found that approximately half the respondents encounter diagnostic errors at least monthly. Moreover, compared with the many different safety concerns encountered in practice, physicians perceive diagnostic errors to be more likely to cause serious harm or death compared with other safety concerns.

Standardised patients studies using ‘secret shoppers’ have also been used to estimate the accuracy of diagnosis. In these studies, real or simulated patients with classical presentations of common diseases, like rheumatoid arthritis, asthma, or chronic obstructive pulmonary disease (COPD) are sent anonymously into real practice settings.

The diagnostic error rates reported (13–15%) are very much in line with estimates from the other types of research approaches, and have substantial ‘face’ validity in that the studies are being carried out prospectively in real-world settings. In addition to providing an estimate of diagnostic error rates, this approach offers the unique ability to probe the various factors that promote or detract from optimal diagnosis.

The chief limitation of these studies in regard to estimating error rates is that they are performed under research-directed conditions, presenting a much smaller subset of conditions than would be seen in usual practice. Moreover, in studies seeking to also study the factors relevant to accurate diagnosis, these patients may present with comorbid conditions or contextual complexities that are not representative of typical patients, and case complexity is a major factor in determining diagnostic accuracy.

Second reviews refers to research protocols in the visual subspecialities (eg, radiology, pathology, dermatology) where a second radiologist examines the same films after a first radiologist, or a second pathologist reviews the same biopsy or cytology specimen as another pathologist. These second review studies may be performed under controlled conditions, involving the review of many or mostly abnormal cases. This approach has advantages from a research perspective, but substantially increases the possibility for diagnostic error, which can range from 10% to 50%.

Interestingly, the studies also show that a diagnostician will also disagree with his or her own prior interpretation in a small fraction of cases.

In ‘real world’ situations, the majority of examinations are normal. Under these conditions, a critical abnormality is detected by a second expert reviewer in the range of 2–5%.

Diagnostic testing audits are used to estimate the incidence of error in the clinical laboratory. Thanks to impressive advances in quality control procedures, diagnostic errors in the modern age are rarely the result of an error in the analytical test itself. Most laboratory-related errors now originate from the preanalytical and postanalytical phases, namely issues related to the physician ordering and interpreting the
test result. Overall, it is estimated that laboratory results are misleadingly wrong in 2–4% of cases, and roughly the same error rates are found in diagnostic radiology. The major limitation of diagnostic testing audits is that, although they excel at finding errors in test performance per se, they underestimate the true clinical impact because of the related preanalytical and postanalytical errors occurring outside the walls of the laboratory or radiology department which, typically, escape detection. Lapses in the reliable communication of abnormal test results, for example, is a universal problem, even in systems with advanced electronic medical records. Malpractice claim databases are easily studied and have provided interesting data on diagnostic errors. Typical data (eg, from the Physicians Insurers Institute of America, shown below) reveals that problems related to diagnostic error are the leading cause of paid claims. Essentially identical data has been reported from the Department of Veterans Affairs, Kaiser-Permanente Healthcare System, and CRICO-RMF (CRICO Risk Management Foundation), the self-insurance entity covering the Harvard teaching hospitals, and from studies in the UK. A recent analysis of malpractice cases extracted from the National Practitioner Data Bank over a 25-year period identified 100,249 cases of diagnostic error. Diagnostic error was the most common reason for a claim (29%) and the most costly, averaging $386,849 per claim.

Studying closed claims provides a convenient approach to analysing large numbers of cases, already preclassified as representing diagnostic errors. Limitations of this approach include the fact that only a small subset of errors results in claims, and these, typically, span just a narrow range of clinical conditions, predominated by cases involving cancer or cardiovascular disease in young or middle-aged patients. Moreover, these cases provide only limited opportunities to study the aetiology of diagnostic errors, as the proceedings by their nature are focused more on consequences than causes.

Case Reviews. Many specific symptoms and disease states have been studied using retrospective case reviews, and in each of these the incidence of diagnostic error is unacceptably high. For example, a systematic review of more than 8000 ER patients found a delayed diagnosis of stroke in 9%. Kostopoulou and colleagues reported a systematic review of this literature and identified 21 studies meeting their quality criteria. In one such study, children with asthma experienced a median delay in making the correct diagnosis of more than 3 years, spanning 7 clinic/ER visits. Delayed or wrong diagnosis rates of 10–50% have been identified in studies of coronary artery disease, HIV-associated complications, tuberculosis and a wide range of malignancies. Very similar data has been reported by Gandhi and colleagues, who studied 181 cases of diagnostic error in ambulatory settings (figure 1). Delayed diagnosis of cancer and coronary artery disease were the most commonly identified problems in the study by Gandhi et al., and a growing number of studies have confirmed the frequency of missed opportunities for earlier cancer diagnosis. The point to be made is that whereas diagnostic errors may sometimes reflect encounters with extremely rare diseases or very unusual presentations of common diseases, in many cases, it is a relatively common disease that is mislabelled or missed entirely. A convenience sampling of studies that measured diagnostic error rates in 40 different symptom complexes or diseases is presented in online supplementary appendix table 1.

The chief limitation of case reviews is that they typically rely just on data contained in the medical record, and many diagnostic errors are missed as a result. As an example, Schwartz and colleagues used standardised patients to estimate costs resulting from diagnostic errors in ambulatory practices. Only their knowledge of the actual diagnosis allowed accurate estimates of these costs; virtually none could have been surmised from record reviews alone. An additional limitation is that the medical record, typically, is lacking documentation on what the clinician was thinking at the time the diagnosis was being considered. Finally, random chart reviews are not well suited to measuring incidence rates of diagnostic error because the rate of error is low, thus requiring a large number of reviews.
Voluntary reports. Voluntary error-reporting systems are now in place in most healthcare organisations in the USA, and were expected to provide a reliable way of identifying both minor and major adverse events. When they are used, and especially if follow-up interviews with providers can be obtained, voluntary reporting offers the unique potential to explore both the systemic-related and cognitive aetiologies of diagnostic error. Would you like me to continue with the text?
DISCUSSION
We have proposed that measuring the incidence of diagnostic error in everyday practice is an essential requirement of a comprehensive quality management programme. Of eight methods used to study diagnostic errors, some are more suitable than others in terms of their potential for providing meaningful data on diagnostic error rates (table 1). Analysing closed malpractice claims can provide only relative data on error rates; estimating absolute rates is not possible because so few true errors actually result in claims, and not all claims reflect true errors. Diagnostic testing audits are meaningful only to the extent that audits examine true clinical impact, because many diagnostic testing errors are ultimately detected and corrected, or discarded as meaningless. Of the remaining methods, each has its own advantages and limitations. Moreover, it is probably safe to assume that each of these approaches will underestimate the actual rate of error, and may identify entirely separate cohorts of errors. This problem is compounded by the different definitions and classification systems of diagnostic error that have been applied in these various studies. The goal of knowing the diagnostic error rate in practice may, therefore, require further research to standardise definitions and other methodologic issues, and combining results from several different approaches.

Similarly, the eight different approaches offer differing capabilities in their ability to provide insight regarding the aetiology of diagnostic error. Autopsies and second reviews reveal that an error was made, but not why. Using standardised patients is a particularly powerful way to study these factors, because at least some of the variables (case presentation and complexity, for example) can be controlled. Voluntary reports from physicians also provide unique opportunities to gather insights on the cognitive and system-related factors that might have contributed to the error. With the exception of using standardised patients, none of the approaches are well suited to study the human factor issues, such as distractions, fatigue and workload stress, thought to play dominant roles influencing clinical decision making.

Knowing the incidence of diagnostic error may be less important than being able to measure the likelihood of harm that results. Extrapolating from the Class 1 errors (a major discrepancy that likely leads to the patient’s death) identified at autopsy, Leape, Berwick and Bates estimated that 80 000 deaths per year might be caused by diagnostic error, including both ambulatory and inpatient errors. A recent systematic review of autopsy data concluded that 36 000 deaths a year were due to diagnostic errors in just ICUs alone. These estimates, of course, do not include the many instances of non-fatal injury related to misdiagnosis, events that will be far more numerous, and the many instances where the harm is psychological or financial more than physical. Preventability, however, is a difficult parameter to judge, and these estimates may exaggerate the impact of diagnostic error to the extent that it is overestimated. Both case review studies and closed claims studies find that diagnostic errors are more likely to cause harm than other patient safety problems.

In summary, a wide range of different research approaches have been used to estimate diagnostic error rates, all suggesting that the incidence is unacceptably high. Although true incidence data are lacking, a wide variety of research studies suggest breakdowns in the diagnostic process result in a staggering toll of harm and patient deaths. A recent, authoritative review by the AMA of ambulatory patient safety concerns reached the same conclusion.

What’s missing from these estimates are true incidence data, as typically the denominators are not available or provided. Also missing are the aggregate rates of injury and harm. There is a clear need for additional research to identify better ways to bring diagnostic errors to light. Promising approaches in this regard include the use of trigger tools focusing on diagnostic error, the use of standardised patients, and encouraging both patients and physicians to voluntarily report errors they encounter.

The most fundamental principle of performance improvement is that ‘You can’t fix what you don’t measure’. Efforts to begin addressing diagnostic error must begin with measurement. In no area of patient safety is this need more acute than in trying to identify the true incidence of diagnostic errors, and the harm associated with these events.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES
Narrative review


45 Claessen D, Resar R, Griffin E, et al. Global Trigger Tool shows that adverse events in hospitals may be ten times greater than previously measured. Health Affairs 2011;30:581–89.


64 Leape L, Berwick D, Bates D. Counting deaths from medical errors. JAMA 2002;288:2405.


