Annals of Internal Medicine | ARTICLE

Missed and Delayed Diagnoses in the Ambulatory Setting: A Study of Closed Malpractice Claims

Tejal K. Gandhi, MD, MPH; Allen Kachalia, MD, JD; Eric J. Thomas, MD, MPH; Ann Louise Puopolo, BSN, RN; Catherine Yoon, MS; Troyen A. Brennan, MD, JD; and David M. Studdert, LLB, ScD

Background: Although missed and delayed diagnoses have become an important patient safety concern, they remain largely unstudied, especially in the outpatient setting.

Objective: To develop a framework for investigating missed and delayed diagnoses, advance understanding of their causes, and identify opportunities for prevention.

Design: Retrospective review of 307 closed malpractice claims in which patients alleged a missed or delayed diagnosis in the ambulatory setting.

Setting: 4 malpractice insurance companies.

Measurements: Diagnostic errors associated with adverse outcomes for patients, process breakdowns, and contributing factors.

Results: A total of 181 claims (59%) involved diagnostic errors that harmed patients. Fifty-nine percent (106 of 181) of these errors were associated with serious harm, and 30% (55 of 181) resulted in death. For 59% (106 of 181) of the errors, cancer was the diagnosis involved, chiefly breast (44 claims [24%]) and colorectal (13 claims [7%]) cancer. The most common breakdowns in the diagnostic process were failure to order an appropriate diagnostic test (100 of 181 [55%]), failure to create a proper follow-up plan (81 of 181 [45%]), failure to obtain an adequate history or perform an adequate physical examination (76 of 181 [42%]), and incorrect interpretation of diagnostic tests (67 of 181 [37%]). The leading factors that contributed to the errors were failures in judgment (143 of 181 [79%]), vigilance or memory (106 of 181 [59%]), knowledge (86 of 181 [48%]), patient-related factors (84 of 181 [46%]), and handoffs (36 of 181 [20%]). The median number of process breakdowns and contributing factors per error was 3 for both (interquartile range, 2 to 4).

Limitations: Reviewers were not blinded to the litigation outcomes, and the reliability of the error determination was moderate.

Conclusions: Diagnostic errors that harm patients are typically the result of multiple breakdowns and individual and system factors. Awareness of the most common types of breakdowns and factors could help efforts to identify and prioritize strategies to prevent diagnostic errors.

Ann Intern Med. 2006;145:488-496. www.annals.org

For author affiliations, see end of text.

Missed and delayed diagnoses in the ambulatory setting are an important patient safety problem. The current diagnostic process in health care is complex, chaotic, and vulnerable to failures and breakdowns. For example, one third of women with abnormal results on mammography or Papanicolaou smears do not receive follow-up care that is consistent with well-established guidelines (1, 2), and primary care providers often report delays in reviewing test results (3). Recognition of systemic problems in this area has prompted urgent calls for improvements (4).

However, this type of error remains largely unstudied (4). At least part of the reason is technical: Because omissions characterize missed diagnoses, they are difficult to identify; there is no standard reporting mechanism; and when they are identified, documentation in medical records is usually insufficiently detailed to support detailed causal analyses. The result is a relatively thin evidence base from which to launch efforts to combat diagnostic errors. Moreover, conceptions of the problem tend to remain rooted in the notion of physicians failing to be vigilant or up-to-date. This is a less nuanced view of error causation than careful analysis of other major patient safety problems, such as medication errors (5, 6), has revealed.

Several considerations highlight malpractice claims as a potentially rich source of information about missed and delayed diagnoses. First, misdiagnosis is a common allegation. Over the past decade, lawsuits alleging negligent misdiagnoses have become the most prevalent type of claim in the United States (7, 8). Second, diagnostic breakdowns that lead to claims tend to be associated with especially severe outcomes. Third, relatively thorough documentation on what happened is available in malpractice insurers' claim files. In addition to the medical record, these files include depositions, expert opinions, and sometimes the results of internal investigations.

Previous attempts to use data from malpractice claims to study patient safety have had various methodologic constraints, including small sample size (9, 10), a focus on single insurers (11) or verdicts (9, 10) (which constitute <10% of claims), limited information on the claims (8–11), reliance on internal case review by insurers rather than by independent experts (8, 11), and a general absence of

See also:

Print
Editors' Notes 489
Editorial comment. 547
Summary for Patients. I-12

Web-Only
Appendix
Appendix Tables
Appendix Figures
Conversion of tables and figures into slides

robust frameworks for classifying types and causes of failures. To address these issues, we analyzed data from closed malpractice claims at 4 liability insurance companies. Our goals were to develop a framework for investigating missed and delayed diagnoses, advance understanding of their causes, and identify opportunities for prevention.

Context

Efforts to reduce medical errors and improve patient safety have not generally addressed errors in diagnosis. As with treatment, diagnosis involves complex, fragmented processes within health care systems that are vulnerable to failures and breakdowns.

Contributions

The authors reviewed malpractice claims alleging injury from a missed or delayed diagnosis. In 181 cases in which there was a high likelihood that error led to the missed diagnosis, the authors analyzed where the diagnostic process broke down and why. The most common missed diagnosis was cancer, and the most common breakdowns were failure to order appropriate tests and inadequate follow-up of test results. A median of 3 process breakdowns occurred per error, and 2 or more clinicians were involved in 43% of cases.

Cautions

The study relied on malpractice claims, which are not representative of all diagnostic errors that occur. There was only moderate agreement among the authors in their subjective judgments about errors and their causes.

Implications

Like other medical errors, diagnostic errors are multifactorial. They arise from multiple process breakdowns, usually involving multiple providers. The results highlight the challenge of finding effective ways to reduce diagnostic errors as a component of improving health care quality.

—The Editors

METHODS

Study Sites

Four malpractice insurance companies based in 3 regions (northeastern, southwestern, and western United States) participated in the study. Collectively, the participating companies insured approximately 21 000 physicians, 46 acute care hospitals (20 academic and 26 nonacademic), and 390 outpatient facilities, including a wide variety of primary care and outpatient specialty practices. The ethics review boards at the investigators' institutions and at each review site approved the study.

Claims Sample

Data were extracted from random samples of closed claim files from each insurer. A claim is classified as closed when it has been dropped, dismissed, paid by settlement, or resolved by verdict. The claim file is the repository of information accumulated by the insurer during the life of a claim. It captures a wide variety of data, including the statement of claim, depositions, interrogatories, and other litigation documents; reports of internal investigations, such as risk management evaluations and sometimes root-cause analyses; expert opinions from both sides; medical reports detailing the plaintiff's preevent and postevent condition; and, while the claim is open, medical records pertaining to the episode of care at issue. We reacquired the relevant medical records for sampled claims.

Following previous studies, we defined a claim as a written demand for compensation for medical injury (12, 13). Claims involving missed or delayed diagnoses were defined as those alleging an error in diagnosis or testing that caused a delay in appropriate treatment or a failure to act or follow up on results of diagnostic tests. We excluded allegations related to pregnancy and those pertaining to care rendered solely in the inpatient setting.

We reviewed 429 diagnostic claims alleging injury due to missed or delayed diagnoses. Insurers contributed to the study sample in proportion to their annual claims volume (Appendix, available at www.annals.org). The claims were divided into 2 main categories based on the primary setting of the outpatient care involved in the allegation: the emergency department (122 claims) and all other locations (for example, physician's office, ambulatory surgery, pathology laboratory, or radiology suites) (307 claims). The latter group, which we call *ambulatory claims*, is the focus of this analysis.

Study Instruments and Claim File Review

Physicians who were board-certified attendings, fellows, or third-year residents in internal medicine reviewed sampled claim files at the insurers' offices or insured facilities. Physician-investigators trained the reviewers in the content of claim files, use of the study instruments, and confidentiality procedures in 1-day sessions at each site. The reviewers also used a detailed manual. Reviews took on average 1.4 hours per file. To test review reliability, a second reviewer reviewed a random selection of 10% (42 of 429) of the files. Thirty-three of the 307 ambulatory claims that are the focus of this analysis were included in the random blinded re-review. A sequence of 4 instruments guided the review. For all claims, insurance staff recorded administrative details of the case (**Appendix Figure 1**, available at www.annals.org), and clinical reviewers recorded details of the adverse outcome the patient experienced, if any (**Appendix Figure 2**, available at www.annals.org). Reviewers scored adverse outcomes on a 9-point severity scale ranging from emotional injury only to death. This scale was developed by the National Association of Insurance Commissioners (14) and has been used in previous research (15). If the patient had multiple adverse outcomes, reviewers scored the most severe outcome. To simplify presentation of our results, we grouped scores on

this scale into 5 categories (emotional, minor, significant, major, and death).

Next, reviewers considered the potential role of a series of contributing factors (**Appendix Figure 3**, available at www.annals.org) in causing the adverse outcome. The factors covered cognitive-, system-, and patient-related causes that were related to the episode of care as a whole, not to particular steps in the diagnostic process. The factors were selected on the basis of a review of the patient safety literature performed in 2001 by 5 of the authors in consultation with physician-collaborators from surgery and obstetrics and gynecology.

Reviewers then judged, in light of available information and their decisions about contributing factors, whether the adverse outcome was due to diagnostic error. We used the Institute of Medicine's definition of error, namely, "the failure of a planned action to be completed as intended (i.e. error of execution) or the use of a wrong plan to achieve an aim (i.e. error of planning)" (16). Reviewers recorded their judgment on a 6-point confidence scale ranging from "1. Little or no evidence that adverse outcome resulted from error/errors" to "6. Virtually certain evidence that adverse outcome resulted from error/errors." Claims that scored 4 ("More likely than not that adverse outcome resulted from error/errors; more than 50-50 but a close call") or higher were classified as having an error. The confidence scale and cutoff point were adapted from instruments used in previous studies of medical injury (15, 17).

Reviewers were not blinded to the litigation outcomes but were instructed to ignore them and rely on their own clinical judgment in making decisions about errors. Training sessions stressed that the study definition of error is not synonymous with the legal definition of negligence and that a mix of factors extrinsic to merit influence whether claims are paid during litigation.

Finally, for the subset of claims judged to involve errors, reviewers completed an additional form (**Appendix Figure 4**, available at www.annals.org) that collected additional clinical information about the missed diagnosis. Specifically, reviewers considered a defined sequence of diagnostic steps (for example, history and physical examination, test ordering, and creation of a follow-up plan) and were asked to grade their confidence that a breakdown had occurred at each step (5-point Likert scale ranging from "highly unlikely" to "highly likely"). If a breakdown was judged to have been at least "somewhat likely" (score of $\geq$3), the form elicited additional information on the particular breakdown, including a non–mutually exclusive list of reasons for the breakdown.

Statistical Analysis

The primary unit of analysis was the sequence of care in claims judged to involve a diagnostic error that led to an adverse outcome. For ease of exposition, we henceforth refer to such sequences as errors. The hand-filled data forms were electronically entered and verified by a professional data entry vendor and sent to the Harvard School of Public Health for analysis. Additional validity checks and data cleaning were performed by study programmers. Analyses were conducted by using SAS, version 8.2 (SAS Institute, Cary, North Carolina) and Stata SE, version 8.0 (Stata Corp., College Station, Texas). We examined characteristics of the claims, patients, and injuries in our sample and the frequency of the various contributing factors. We compared characteristics of error subgroups using Pearson chi-square tests and used percentage agreement and κ scores (18) to measure interrater reliability of the injury and error determinations.

Role of the Funding Sources

This study was funded by grants from the Agency for Healthcare Research and Quality (HS011886-03) and the Harvard Risk Management Foundation. These organizations did not play a role in the design, conduct, or analysis of the study, and the decision to submit the manuscript for publication was that of the authors.

RESULTS

The 307 diagnosis-related ambulatory claims closed between 1984 and 2004. Eighty-five percent (262 of 307) of the alleged errors occurred in 1990 or later, and 80% (245 of 307 claims) closed in 1997 or later. In 2% (7 of 307) of claims, no adverse outcome or change in the patient's clinical course was evident; in 3% (9 of 307), the reviewer was unable to judge the severity of the adverse outcome from the information available; and in 36% (110 of 307), the claim was judged not to involve a diagnostic error. The remaining group of 181 claims, 59% (181 of 307) of the sample, were judged to involve diagnostic errors that led to adverse outcomes. This group of errors is the focus of further analyses.

In 40 of the 42 re-reviewed claims, reviewers agreed about whether an adverse outcome had occurred (95% agreement). The reliability of the determination of whether an error had occurred (a score $<$4 vs. $\geq$4 on the confidence scale) was moderate (72% agreement; $\kappa = 0.42$ [95% CI, -0.05 to 0.66]).

Errors and Diagnoses

Fifty-nine percent (106 of 181) of the errors were associated with significant or major physical adverse outcomes and 30% (55 of 181) were associated with death (**Table 1**). For 59% (106 of 181) of errors, cancer was the diagnosis missed, chiefly breast (44 of 181 [24%]), colorectal (13 of 181 [7%]), and skin (8 of 181 [4%]) cancer. The next most commonly missed diagnoses were infections (9 of 181 [5%]), fractures (8 of 181 [4%]), and myocardial infarctions (7 of 181 [4%]). Most errors occurred in physicians' offices (154 of 181 [85%]), and primary care physicians were the providers most commonly involved (76 of 181 [42%]). The mean interval between when diagnoses should have been made (that is, in the absence of error)

Table 1. **Characteristics of Patients, Involved Clinicians, Adverse Outcomes, and Missed or Delayed Diagnoses among 181 Diagnostic Errors in Ambulatory Care**

Characteristics	Errors, *n* (%)
Female patient	110 (61)
Patient age (mean, 44 y [SD, 17])	
<1 y	5 (3)
1–18 y	9 (5)
18–34 y	34 (19)
35–49 y	64 (35)
50–64 y	50 (28)
>64 y	19 (10)
Health insurance (*n* = 117)	
Private	103 (88)
Medicaid	3 (3)
Uninsured	7 (6)
Medicare	1 (1)
Other	3 (3)
Clinicians involved in error*	
Specialty	
Primary care†	88 (49)
Radiology	31 (17)
General surgery	23 (13)
Pathology	13 (7)
Physician's assistant	13 (7)
Registered nurse or nurse practitioner	14 (8)
Trainee (resident, fellow, or intern)	20 (11)
Setting	
Physician's office	154 (85)
Ambulatory surgery facility	8 (4)
Pathology or clinical laboratory	8 (4)
Radiology suite	5 (3)
Other	6 (3)
Missed or delayed diagnosis	
Cancer	106 (59)
Breast	44 (24)
Colorectal	13 (7)
Skin	8 (4)
Hematologic	7 (4)
Gynecologic	7 (4)
Lung	6 (3)
Brain	5 (3)
Prostate	5 (3)
Liver or gastric	2 (1)
Other	9 (5)
Infection	9 (5)
Fracture	8 (4)
Myocardial infarction	7 (4)
Embolism	6 (3)
Appendicitis	5 (3)
Cerebral vascular disease	4 (2)
Other neurologic condition	4 (2)
Aneurysm	3 (2)
Other cardiac condition	3 (2)
Gynecologic disease (noncancer)	3 (2)
Endocrine disorder	3 (2)
Birth defect	3 (2)
Ophthalmologic disease	3 (2)
Peripheral vascular disease	2 (1)
Abdominal disease	2 (1)
Psychiatric illness	2 (1)
Other	8 (4)
Adverse outcome‡	
Psychological or emotional	9 (5)
Minor physical	11 (6)
Significant physical	72 (40)
Major physical	34 (19)
Death	55 (30)

* Percentages do not sum to 100% because multiple providers were involved in some errors.
† Includes 12 pediatricians, therefore pediatrics accounts for an absolute 7% of the 49%.
‡ Categories correspond to the following scores on the National Association of Insurance Commissioners' 9-point severity scale: psychiatric or emotional (score 1), minor physical (scores 2 and 3), significant physical (scores 4 to 6), major physical (scores 7 and 8), and death (score 9). For details, see Appendix Figure 2, available at www.annals.org.

and when they actually were made was 465 days (SD, 571 days), and the median was 303 days (interquartile range, 36 to 681 days). **Appendix Table 1** (available at www.annals.org) outlines several examples of missed and delayed diagnoses in the study sample.

Breakdowns in the Diagnostic Process

The leading breakdown points in the diagnostic process were failure to order an appropriate diagnostic test (100 of 181 [55%]), failure to create a proper follow-up plan (81 of 181 [45%]), failure to obtain an adequate history or to perform an adequate physical examination (76 of 181 [42%]), and incorrect interpretation of a diagnostic test (67 of 181 [37%]) (**Table 2**). Missed cancer diagnoses were significantly more likely to involve diagnostic tests being performed incorrectly (14 of 106 [13%] vs. 1 of 75 [1%]; $P = 0.004$) and being interpreted incorrectly (49 of 106 [46%] vs. 18 of 75 [24%]; $P = 0.002$), whereas missed noncancer diagnoses were significantly more likely to involve delays by patients in seeking care (4 of 106 [4%] vs. 12 of 75 [16%]; $P = 0.004$), inadequate history or physical examination (24 of 106 [23%] vs. 52 of 75 [69%]; $P < 0.001$), and failure to refer (19 of 106 [18%] vs. 28 of 75 [37%]; $P = 0.003$).

Table 3 details the 4 most common process breakdowns. Among failures to order diagnostic tests, most tests involved were from the imaging (45 of 100 [45%]) or other test (50 of 100 [50%]) categories. With respect to the specific tests involved, biopsies were most frequently at issue (25 of 100 [25%]), followed by computed tomography scans, mammography, ultrasonography, and colonoscopy (11 of 100 for each [11%]). The most common explanation for the failures to order was that the physician seemed to lack knowledge of the appropriate test in the clinical circumstances. Among misinterpretations of test results, 27% (18 of 67) related to results on mammography and 15% (10 of 67) related to results on radiography. The main reasons for inadequate follow-up plans were that the physician did not think follow-up was necessary (32 of 81 [40%]), selected an inappropriate follow-up interval (29 of

Table 2. **Breakdown Points in the Diagnostic Process**

Process Breakdown	All Missed Diagnoses (n = 181), n (%)	Missed Cancer Diagnoses (n = 106), n (%)	Missed Noncancer Diagnoses (n = 75), n (%)	P Value*
Initial delay by the patient in seeking care	16 (9)	4 (4)	12 (16)	0.004
Failure to obtain adequate medical history or physical examination	76 (42)	24 (23)	52 (69)	<0.001
Failure to order appropriate diagnostic or laboratory tests	100 (55)	63 (59)	37 (49)	0.178
Adequate diagnostic or laboratory tests ordered but not performed	17 (9)	10 (9)	7 (9)	0.98
Diagnostic or laboratory tests performed incorrectly	15 (8)	14 (13)	1 (1)	0.004
Incorrect interpretation of diagnostic or laboratory tests	67 (37)	49 (46)	18 (24)	0.002
Responsible provider did not receive diagnostic or laboratory test results	23 (13)	17 (16)	6 (8)	0.110
Diagnostic or laboratory test results were not transmitted to patient	22 (12)	15 (14)	7 (9)	0.33
Inappropriate or inadequate follow-up plan	81 (45)	51 (48)	30 (40)	0.28
Failure to refer	47 (26)	19 (18)	28 (37)	0.003
Failure of a requested referral to occur	9 (5)	8 (8)	1 (1)	0.058
Failure of the referred-to clinician to convey relevant results to the referring clinician	3 (2)	2 (2)	1 (1)	0.77
Patient nonadherence to the follow-up plan	31 (17)	21 (20)	10 (13)	0.25

* Pearson chi-square test for differences between cancer and noncancer categories.

81 [36%]), or did not document the plan correctly (22 of 81 [27%]).

Contributing Factors

The leading factors that contibuted to errors were failures in judgment (143 of 181 [79%]), vigilance or memory (106 of 181 [59%]), knowledge (86 of 181 [48%]), patient-related factors (84 of 181 [46%]), and handoffs (36 of 181 [20%]) (**Table 4**). The patient-related factors included nonadherence (40 of 181 [22%]), atypical clinical presentation (28 of 181 [15%]), and complicated medical history (18 of 181 [10%]). There were no significant differences in the prevalence of the various contributing factors when missed cancer and missed noncancer diagnoses were compared, with the exception of lack of supervision, which was significantly more likely to have occurred in missed noncancer diagnoses (10 of 75 [13%] vs. 5 of 106 [5%]; $P = 0.038$). The higher prevalence of lack of supervision as a contributing factor to missed noncancer diagnoses was accompanied by greater involvement of trainees in these cases (13 of 75 [17%] vs. 7 of 106 [7%]; $P = 0.023$).

Multifactorial Nature of Missed or Delayed Diagnoses

The diagnostic errors were complex and frequently involved multiple process breakdowns, contributing factors, and contributing clinicians (**Table 5**). In 43% (78 of 181) of errors, 2 or more clinicians contributed to the missed diagnosis, and in 16% (29 of 181), 3 or more clinicians contributed. There was a median of 3 (interquartile range, 2 to 4) process breakdowns per error; 54% (97 of 181) of errors had 3 or more process breakdowns and 29% (52 of 181) had 4 or more. Thirty-five diagnostic errors (35 of 181 [19%]) involved a breakdown at only 1 point in the care process (*single-point breakdowns*). Twenty-four (24 of 35 [69%]) of these single-point breakdowns were either failures to order tests ($n = 10$) or incorrect interpretations of tests ($n = 14$); only 3 involved trainees.

The median number of contributing factors involved in diagnostic errors was 3 (interquartile range, 2 to 4); 59% (107 of 181) had 3 or more contributing factors, 27% (48 of 181) had 4 or more, and 13% (23 of 181) had 5 or more. Thus, although virtually all diagnostic errors were linked to cognitive factors, especially judgment errors, cognitive factors operated alone in a minority of cases. They were usually accompanied by communication factors, patient-related factors, or other system factors.

Specifically, 36% (66 of 181) of errors involved cognitive factors alone, 16% (29 of 181) involved judgment or vigilance and memory factors alone, and 9% (16 of 181) involved only judgment factors. The likelihood that cognitive factors alone led to the error was significantly lower among errors that involved an inadequate medical history or physical examination (21 of 66 [32%] vs. 55 of 115 [48%]; $P = 0.036$), lack of receipt of ordered tests by the responsible provider (3 of 66 [5%] vs. 20 of 115 [17%]; $P = 0.013$), and inappropriate or inadequate follow-up planning (21 of 66 [32%] vs. 60 of 115 [52%]; $P = 0.008$). In other words, the role of communication and other systems factors was especially prominent in these 3 types of breakdowns.

DISCUSSION

Our study of closed malpractice claims identified a group of missed diagnoses in the ambulatory setting that were associated with dire outcomes for patients. Over half of the missed diagnoses were cancer, primarily breast and colorectal cancer; no other diagnosis accounted for more

than 5% of the sample. The main breakdowns in the diagnostic process were failure to order appropriate diagnostic tests, inappropriate or inadequate follow-up planning, failure to obtain an adequate medical history or perform an adequate physical examination, and incorrect interpretation of diagnostic test results. Cognitive factors, patient-related factors, and handoffs were the most prominent contributing factors overall. However, relatively few diagnostic errors could be linked to single-point breakdowns or lone contributing factors. Most missed diagnoses involved manifold breakdowns and a potent combination of individual and system factors. The resultant delays tended to be long, setting diagnosis back more than 1 year on average.

The threshold question of what constitutes a harmful diagnostic error is extremely challenging. Indeed, the nebulous nature of missed diagnoses probably helps to explain why patient safety research in this area has lagged. Our approach was 2-dimensional. We considered the potential role of a range of contributing factors drawn from the fields of human factors and systems analysis (16, 19) in causing the patient's injury. We also disaggregated the diagnostic process into discrete steps and then examined the prevalence of problems within particular steps. Trained physician-reviewers had the benefit of information from the claim file and from the medical record in deciding whether diagnoses were missed. Even so, agreement among reviewers was only fair, highlighting how difficult and subjective clinical judgments regarding missed diagnoses are.

Table 3. **Details of the 4 Most Frequent Process Breakdowns***

Process Breakdown	Leading Reasons for Breakdown	Value, *n*	Within Process Breakdown Category, %
Failure to order appropriate diagnostic or laboratory tests (n = 100)	Provider lacked knowledge of appropriate test	37	37
	Poor documentation	10	10
	Failure of communication	15	15
	Among providers	7	7
	Between provider and patient	8	8
	Patient did not schedule or keep appointment	9	9
	Patient declined tests	5	5
Type of test not ordered	Imaging	45	45
	Computed tomography scanning	11	11
	Mammography	11	11
	Ultrasonography	11	11
	Blood	14	14
	Other†	50	50
	Biopsy	25	25
	Colonoscopy or flexible sigmoidoscopy	11	11
Inappropriate or inadequate follow-up plan (n = 81)	Provider did not think follow-up was necessary	32	40
	Provider selected inappropriate follow-up interval	29	36
	Follow-up plan not documented	22	27
	Follow-up appointment not scheduled	20	25
	Miscommunication between patient and provider	20	25
	Miscommunication between providers	10	12
Failure to obtain adequate medical history or physical examination (n = 76)	Incomplete physical examination	39	50
	Failure to elicit relevant information	35	46
	Poor documentation	20	26
	Patient provided inaccurate history	13	17
Incorrect interpretation of diagnostic or laboratory tests (n = 67)	Error in clinical judgment	52	78
	Failure of communication among providers	5	7
	Inexperience	3	5
	Whose misinterpretation?		
	Radiologist	27	40
	Primary care physician	19	28
	Pathologist	10	15
Type of test interpreted incorrectly	Imaging	36	54
	Mammography	18	27
	Radiography	10	15
	Ultrasonography	5	7
	Blood	9	13
	Other‡	21	31
	Biopsy	8	12

* Category and subcategory totals may exceed 100% because of non–mutually exclusive reasons and multiple tests. Only leading reasons and tests are presented.
† The tests not shown in this category are echocardiography (n = 3), cardiac catheterization (n = 3), electrocardiography (n = 2), treadmill (n = 2), urinalysis (n = 1), esophagogastroduodenoscopy/barium swallow (n = 1), colposcopy (n = 1), urine pregnancy (n = 1), and joint aspirate (n = 1).
‡ The tests not shown in this category are electrocardiography (n = 4), urinalysis (n = 1), echocardiography (n = 1), stool guaiac (n = 2), Papanicolaou smear (n = 3), cystoscopy (n = 1), and treadmill (n = 1).

Table 4. **Factors That Contributed to Missed or Delayed Diagnoses**

Variable	Value, *n* (%)*
Provider- and systems-related factors	
Cognitive factors	179 (99)
Judgment	143 (79)
Vigilance or memory	106 (59)
Lack of knowledge	86 (48)
Communication factors	55 (30)
Handoffs	36 (20)
Failure to establish clear lines of responsibility	17 (9)
Conflict	3 (2)
Some other failure of communication	16 (9)
Other systems factors	31 (17)
Lack of supervision	15 (8)
Workload	12 (7)
Interruptions	6 (3)
Technology problems	5 (3)
Fatigue	1 (1)
Patient-related factors†	84 (46)
Nonadherence	40 (22)
Atypical presentation	28 (15)
Complicated medical history	18 (10)
Information about medical history of poor quality	8 (4)
Other	15 (8)

* Percentages were calculated by using the denominator of 181 missed or delayed diagnoses.
† Sum of specific patient-related factors exceeds total percentage because some errors had multiple contributing factors.

Troublesome Diagnoses

The prevalence of missed cancer diagnoses in our sample is consistent with previous recognition of this problem as a major quality concern for ambulatory care (8). Missed cancer diagnoses were more likely than other missed diagnoses to involve errors in the performance and interpretation of tests. Lack of adherence to guidelines for cancer screening and test ordering is a well-recognized problem, and there are ongoing efforts to address it (1, 20). Our data did not allow assessment of the extent to which guideline adherence would have prevented errors.

The next most commonly missed diagnoses were infections, fractures, and myocardial infarctions, although there was also a broad range of other missed diagnoses. Primary care physicians were centrally involved in most of these errors. Most primary care physicians work under considerable time pressure, handle a wide range of clinical problems, and treat predominantly healthy patients. They practice in a health care system in which test results are not easily tracked, patients are sometimes poor informants, multiple handoffs exist, and information gaps are the norm. For even the most stellar practitioners, clinical processes and judgments are bound to fail occasionally under such circumstances.

Table 5. **Distribution of Contributing Clinicians, Process Breakdowns, and Contributing Factors***

Number per Error	Contributing Clinicians, %	Process Breakdowns, %	Contributing Factors, %
1	57	19	14
≥2	43	81	86
≥3	16	54	59
≥4	6	29	27
≥5	2	11	13

* The total number of errors was 181.

Cognitive and Other Factors

Although there were several statistically significant differences in the patterns of process breakdowns between missed cancer diagnoses, the patterns of contributing factors were remarkably similar. This suggests that although vulnerable points may vary according to the idiosyncratic pathways of different diagnoses, a core set of human and system vulnerabilities permeates diagnostic work. For example, relying exclusively on a physician's knowledge or memory to ensure that the correct test is ordered, practicing in a clinic with poor internal communication strategies, or having a high prevalence of patients with complex disorders may be risky regardless of the potential diagnosis.

Such cognitive factors as judgment and knowledge were nearly ubiquitous but rarely were the sole contributing factors. Previous research has highlighted the importance of cognitive factors in misdiagnosis and has proposed frameworks for better understanding the underlying causes of such errors (21–23). Detecting these cognitive failures and determining why they occur is difficult, but the design of robust diagnostic processes will probably depend on it.

Follow-Up

Nearly half of all errors involved an inadequate follow-up plan, a failure that was split fairly evenly between situations in which the physician determined that follow-up was not necessary and those in which the need for follow-up was recognized but the wrong interval was selected. Poor documentation, scheduling problems, and miscommunication among providers also played a role. A recent survey (24) of patients in 5 countries found that 8% to 20% of adults who were treated as outpatients did not receive their test results, and 9% to 15% received incorrect results or delayed notification of abnormal results. In addition, in our study, 56% (45 of 81) of the errors that occurred during follow-up also had patient-related factors that contributed to the poor outcome, a finding that highlights the need for risk reduction initiatives that address potential breakdowns on both sides of the patient–physician relationship during follow-up.

Interventions

In general, our findings reinforce the need for systems interventions that mitigate the potential impact of cognitive errors by reducing reliance on memory, forcing consideration of alternative diagnostic plans or second opinions, and providing clinical decision support systems (21, 22). However, cognitive factors rarely appear alone, so interventions must also address other contributing factors, such as handoffs and communication. To be clinically use-

ful and maximally effective, such interventions must be easy to use and well integrated into clinicians' workflows (25). For example, incorporating clinical decision support into the electronic medical record of a patient with breast symptoms should help guard against failures in history taking or test ordering by bolstering physicians' knowledge and reducing their reliance on vigilance and memory.

An important feature of interventions is that their use should not rely on voluntary decisions to access them, because physicians are often unaware of their need for help (26). Rather, the use of an intervention should be automatically triggered by certain predetermined and explicit characteristics of the clinical encounter. For example, strategies to combat misinterpretation could mandate second reviews of test results in designated circumstances or require rapid expert reviews when physicians interpret test results outside of their areas of expertise.

After follow-up plans are made, improvements in scheduling procedures, tickler systems, and test result tracking systems could help keep patients and physicians on track. Nationwide attention is being given to the issue of follow-up (27, 28). Research and quality improvement efforts that equip physicians with tools to reliably perform follow-up must be a priority.

Selecting just 1 of the interventions we have outlined may not be sufficient. The multifactorial and complex nature of diagnostic errors suggests that meaningful reductions will require prevention strategies that target multiple levels in the diagnostic process and multiple contributing factors. Nevertheless, the resource constraints most health care institutions face demand priorities. Attention to the 3 vulnerable points we identified—ordering decisions, test interpretation, and follow-up planning—is a useful starting point and promises high yields, especially in reducing the number of missed cancer diagnoses.

Study Limitations

Our study has several limitations. First, unlike prospective observational studies or root-cause analyses, retrospective review of records, even the detailed records found in malpractice claim files, will miss certain breakdowns (for example, patient nonadherence) and contributing factors (for example, fatigue and workload), unless they emerged as issues during litigation. This measurement problem means that prevalence findings for such estimates will be lower bounds, and the multifactorial causality we observed probably understates the true complexity of diagnostic errors.

An additional measurement problem relates to the process breakdowns. Although reviewers considered breakdowns as independent events, in some situations breakdowns may have been prompted or influenced by earlier breakdowns in the diagnostic sequence. For example, a failure to order appropriate tests may stem from oversights in the physical examination. Such interdependence would tend to inflate the frequency of some breakdowns.

Second, awareness of the litigation outcome may have biased reviewers toward finding errors in claims that received compensation and vice versa (29, 30). Several factors militate against this bias: Reviewers were instructed to ignore the litigation outcome; physicians, who as a group tend to be skeptical of the malpractice system, may have been disinclined to credit the system's findings; and, in fact, one quarter of error judgments diverged from the litigation outcomes.

Third, the reliability of the error determination was not high, and the CIs around the κ score are statistically compatible with poor or marginal reliability. Diagnostic errors are one of the most difficult types of errors to detect reliably (31). Although reviewers may have included some episodes of care in the study sample that were not "true" errors, and may have overlooked some that were, we know of no reason why the characteristics of such false-positive results and false-negative results would differ systematically from those we analyzed.

Fourth, malpractice claims data generally, and in our sample in particular, have several other biases. Severe injuries and younger patients are overrepresented in the subset of patients with medical injuries that trigger litigation (32, 33). It is possible that the factors that lead to errors in litigated cases may differ systematically from the factors that lead to errors in nonlitigated cases, although we know of no reason why they would. In addition, outpatient clinics associated with teaching hospitals are overrepresented in our sample, so the diagnostic errors we identified may not be generalizable outside this setting.

Conclusions

Our findings highlight the complexity of diagnostic errors in ambulatory care. Just as Reason's "Swiss cheese" model of accident causation suggests (19) diagnostic errors that harm patients seem to result from the alignment of multiple breakdowns, which in turn stem from a confluence of contributing factors. The task of effecting meaningful improvements to the diagnostic process—with its numerous clinical steps, stretched across multiple providers and months or years, and the heavy reliance on patient initiative—looms as a formidable challenge. The prospects for "silver bullets" in this area seem remote.

Are meaningful gains achievable in the short to medium term? The answer will probably turn on whether simple interventions that target 2 or 3 critical breakdown points are sufficient to disrupt the causal chain or whether interventions at a wider range of points are necessary to avert harm. In this sense, our findings are humbling, and they underscore the need for continuing efforts to develop the "basic science" of error prevention in medicine (34), which remains in its infancy.

From Brigham and Women's Hospital and Harvard School of Public Health, Boston, Massachusetts; University of Texas Health Science Center, Houston, Texas; and the Harvard Risk Management Foundation, Cambridge, Massachusetts.

Note: Drs. Gandhi, Kachalia, and Studdert had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Grant Support: This study was supported by grants from the Agency for Healthcare Research and Quality (HS011886-03) and the Harvard Risk Management Foundation. Dr. Studdert was also supported by the Agency for Healthcare Research and Quality (KO2HS11285).

Potential Financial Conflicts of Interest: *Employment:* T.A. Brennan (Aetna); *Stock ownership or options (other than mutual funds):* T.A. Brennan (Aetna); *Expert testimony:* T.A. Brennan.

Requests for Single Reprints: David M. Studdert, LLB, ScD, Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115; e-mail, studdert@hsph.harvard.edu.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Haas JS, Cook EF, Puopolo AL, Burstin HR, Brennan TA.** Differences in the quality of care for women with an abnormal mammogram or breast complaint. J Gen Intern Med. 2000;15:321-8. [PMID: 10840267]

2. **Marcus AC, Crane LA, Kaplan CP, Reading AE, Savage E, Gunning J, et al.** Improving adherence to screening follow-up among women with abnormal Pap smears: results from a large clinic-based trial of three intervention strategies. Med Care. 1992;30:216-30. [PMID: 1538610]

3. **Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW.** "I wish I had seen this test result earlier!": dissatisfaction with test result management systems in primary care. Arch Intern Med. 2004;164:2223-8. [PMID: 15534158]

4. **Margo CE.** A pilot study in ophthalmology of inter-rater reliability in classifying diagnostic errors: an underinvestigated area of medical error. Qual Saf Health Care. 2003;12:416-20. [PMID: 14645756]

5. **Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al.** Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA. 1995;274:29-34. [PMID: 7791255]

6. **Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, et al.** Systems analysis of adverse drug events. ADE Prevention Study Group. JAMA. 1995;274:35-43. [PMID: 7791256]

7. **Chandra A, Nundy S, Seabury SA.** The growth of physician medical malpractice payments: evidence from the national practitioner data bank. Health Aff (Millwood). 2005;W5240-9. [PMID: 15928255]

8. **Phillips RL Jr, Bartholomew LA, Dovey SM, Fryer GE Jr, Miyoshi TJ, Green LA.** Learning from malpractice claims about negligent, adverse events in primary care in the United States. Qual Saf Health Care. 2004;13:121-6. [PMID: 15069219]

9. **Kern KA.** Medicolegal analysis of bile duct injury during open cholecystectomy and abdominal surgery. Am J Surg. 1994;168:217-22. [PMID: 8080055]

10. **Kern KA.** Malpractice litigation involving laparoscopic cholecystectomy. Cost, cause, and consequences. Arch Surg. 1997;132:392-8. [PMID: 9108760]

11. **Kravitz RL, Rolph JE, McGuigan K.** Malpractice claims data as a quality improvement tool. I. Epidemiology of error in four specialties. JAMA. 1991;266:2087-92. [PMID: 1920696]

12. **Weiler PC, Hiatt HH, Newhouse JP, Johnson WG, Brennan T, Leape LL.** A Measure of Malpractice: Medical Injury, Malpractice Litigation, and Patient Compensation. Cambridge, MA: Harvard Univ Pr; 1993.

13. **Studdert DM, Brennan TA, Thomas EJ.** Beyond dead reckoning: measures of medical injury burden, malpractice litigation, and alternative compensation models from Utah and Colorado. Indiana Law Rev. 2000;33:1643-86.

14. **Sowka M, ed.** National Association of Insurance Commissioners. Malpractice Claims: Final Compilation. Brookfield, WI: National Association of Insurance Commissioners; 1980.

15. **Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al.** Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care. 2000;38:261-71. [PMID: 10718351]

16. **Kohn LT, Corrigan JM, Donaldson MS.** To Err Is Human. Building a Safer Health System. Institute of Medicine. Washington, DC: National Academy Pr; 1999.

17. **Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al.** Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med. 1991;324:370-6. [PMID: 1987460]

18. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74. [PMID: 843571]

19. **Reason J.** Managing the Risks of Organizational Accidents. Brookfield, VT: Ashgate; 1997.

20. **Harvard Risk Management Foundation.** Breast care management algorithm: improving breast patient safety. Accessed at www.rmf.harvard.edu/files/flash/breast-care-algorithm/index.htm on 9 August 2006.

21. **Croskerry P.** The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78:775-80. [PMID: 12915363]

22. **Graber M, Gordon R, Franklin N.** Reducing diagnostic errors in medicine: what's the goal? Acad Med. 2002;77:981-92. [PMID: 12377672]

23. **Elstein AS, Schwarz A.** Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. BMJ. 2002;324:729-32. [PMID: 11909793]

24. **Schoen C, Osborn R, Huynh PT, Doty M, Davis K, Zapert K, et al.** Primary care and health system performance: adults' experiences in five countries. Health Aff (Millwood). 2004;Suppl:W4-487-503. [PMID: 15513956]

25. **Kawamoto K, Houlihan CA, Balas EA, Lobach DF.** Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330:765. [PMID: 15767266]

26. **Friedman CP, Gatti GG, Franz TM, Murphy GC, Wolf FM, Heckerling PS, et al.** Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. J Gen Intern Med. 2005;20:334-9. [PMID: 15857490]

27. **Joint Commission on Accreditation of Healthcare Organizations.** National Patient Safety Goals. Accessed at www.jointcommission.org/PatientSafety/NationalPatientSafetyGoals/07_amb_obs_npsgs.htm on 9 August 2006.

28. **Gandhi TK.** Fumbled handoffs: one dropped ball after another. Ann Intern Med. 2005;142:352-8. [PMID: 15738454]

29. **Guthrie C, Rachlinski JJ, Wistrich AJ.** Inside the judicial mind. Cornell Law Rev. 2001;86:777-830.

30. **LaBine SJ, LaBine G.** Determinations of negligence and the hindsight bias. Law Hum Behav. 1996;20:501-16.

31. **Studdert DM, Mello MM, Gawande AA, Gandhi TK, Kachalia A, Yoon C, et al.** Claims, errors, and compensation payments in medical malpractice litigation. N Engl J Med. 2006;354:2024-33. [PMID: 16687715]

32. **Burstin HR, Johnson WG, Lipsitz SR, Brennan TA.** Do the poor sue more? A case-control study of malpractice claims and socioeconomic status. JAMA. 1993;270:1697-701. [PMID: 8411499]

33. **Studdert DM, Thomas EJ, Burstin HR, Zbar BI, Orav EJ, Brennan TA.** Negligent care and malpractice claiming behavior in Utah and Colorado. Med Care. 2000;38:250-60. [PMID: 10718350]

34. **Brennan TA, Gawande A, Thomas E, Studdert D.** Accidental deaths, saved lives, and improved quality. N Engl J Med. 2005;353:1405-9. [PMID: 16192489]

Current Author Addresses: Drs. Gandhi and Kachalia and Ms. Yoon: Division of General Internal Medicine, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02114.
Dr. Thomas: University of Texas Medical School at Houston, 6431 Fannin, MSB 1.122, Houston, TX 77030.
Ms. Puopolo: CRICO/RMF, 101 Main Street, Cambridge, MA 02142.
Dr. Brennan: 151 Farmington Avenue, RC5A, Hartford, CT 06156.
Dr. Studdert: Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115.

Author Contributions: Conception and design: T.K. Gandhi, E.J. Thomas, T.A. Brennan, D.M. Studdert.
Analysis and interpretation of the data: T.K. Gandhi, A. Kachalia, C. Yoon, D.M. Studdert.
Drafting of the article: T.K. Gandhi, A. Kachalia, T.A. Brennan, D.M. Studdert.
Critical revision of the article for important intellectual content: T.K. Gandhi, A. Kachalia, E.J. Thomas, T.A. Brennan, D.M. Studdert.
Final approval of the article: T.K. Gandhi, A. Kachalia, E.J. Thomas, D.M. Studdert.
Provision of study materials or patients: A.L. Puopolo.
Statistical expertise: D.M. Studdert.
Obtaining of funding: D.M. Studdert.
Administrative, technical, or logistic support: A. Kachalia, A.L. Puopolo, D.M. Studdert.
Collection and assembly of data: A. Kachalia, A.L. Puopolo, D.M. Studdert.