JAMA Internal Medicine | Original Investigation

Diagnostic Errors in Hospitalized Adults Who Died or Were Transferred to Intensive Care

Andrew D. Auerbach, MD, MPH; Tiffany M. Lee, BA; Colin C. Hubbard, PhD; Sumant R. Ranji, MD; Katie Raffel, MD; Gilmer Valdes, PhD, DABR; John Boscardin, PhD; Anuj K. Dalal, MD; Alyssa Harris, MPH; Ellen Flynn, RN, MBA, JD; Jeffrey L. Schnipper, MD, MPH; for the UPSIDE Research Group

IMPORTANCE Diagnostic errors contribute to patient harm, though few data exist to describe their prevalence or underlying causes among medical inpatients.

OBJECTIVE To determine the prevalence, underlying cause, and harms of diagnostic errors among hospitalized adults transferred to an intensive care unit (ICU) or who died.

DESIGN, SETTING, AND PARTICIPANTS Retrospective cohort study conducted at 29 academic medical centers in the US in a random sample of adults hospitalized with general medical conditions and who were transferred to an ICU, died, or both from January 1 to December 31, 2019. Each record was reviewed by 2 trained clinicians to determine whether a diagnostic error occurred (ie, missed or delayed diagnosis), identify diagnostic process faults, and classify harms. Multivariable models estimated association between process faults and diagnostic error. Opportunity for diagnostic error reduction associated with each fault was estimated using the adjusted proportion attributable fraction (aPAF). Data analysis was performed from April through September 2023.

MAIN OUTCOMES AND MEASURES Whether or not a diagnostic error took place, the frequency of underlying causes of errors, and harms associated with those errors.

RESULTS Of 2428 patient records at 29 hospitals that underwent review (mean [SD] patient age, 63.9 [17.0] years; 1107 [45.6%] female and 1321 male individuals [54.4%]), 550 patients (23.0%; 95% CI, 20.9%-25.3%) had experienced a diagnostic error. Errors were judged to have contributed to temporary harm, permanent harm, or death in 436 patients (17.8%; 95% CI, 15.9%-19.8%); among the 1863 patients who died, diagnostic error was judged to have contributed to death in 121 (6.6%; 95% CI, 5.3%-8.2%). In multivariable models examining process faults associated with any diagnostic error, patient assessment problems (aPAF, 21.4%; 95% CI, 16.4%-26.4%) and problems with test ordering and interpretation (aPAF, 19.9%; 95% CI, 14.7%-25.1%) had the highest opportunity to reduce diagnostic errors; similar ranking was seen in multivariable models examining harmful diagnostic errors.

CONCLUSIONS AND RELEVANCE In this cohort study, diagnostic errors in hospitalized adults who died or were transferred to the ICU were common and associated with patient harm. Problems with choosing and interpreting tests and the processes involved with clinician assessment are high-priority areas for improvement efforts.

Viewpoint and Editor's Note
Multimedia
Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: The members of the UPSIDE Research Group are listed in Supplement 2.

Corresponding Author: Andrew D. Auerbach, MD, MPH, Division of Hospital Medicine, 521 Parnassus Ave, Room 104, Box 0131, San Francisco, CA 94143-0131 (andrew.auerbach@ucsf.edu).

JAMA Intern Med. doi:10.1001/jamainternmed.2023.7347
Published online January 8, 2024.

Diagnostic errors are "the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient."[1(p4)] Many factors contribute to diagnostic errors, but key among them are complex care systems, limited time available to clinicians trying to ascertain a firm diagnosis, and work cultures that impede improvements in diagnostic performance.[2-8]

Diagnostic errors are long recognized components of adverse events in hospitalized patients[9,10] and major factors in closed malpractice claims[11] and are thought to be contributors to trigger events, such as deaths or intensive care unit (ICU) transfers,[12-14] although few past studies used structured approaches to detect diagnostic errors. For example, a recent study of inpatient adverse events did not screen specifically for diagnostic processes and detected diagnostic error in only 10 of nearly 1000 adverse events reviewed.[15] The few studies specifically examining diagnostic errors in medical inpatients have limitations due to differences in the underlying events triggering a review, as well as in review processes used.[16-18]

To address these gaps, we conducted a retrospective multicenter cohort study using a rigorous adjudication process to assess the frequency, underlying causes, and harms of diagnostic errors among adults hospitalized with medical diagnoses between January 1 and December 31, 2019, and who had a trigger event of ICU transfer or death during their stay.

Key Points

Question How often do diagnostic errors happen in adult patients who are transferred to the intensive care unit (ICU) or die in the hospital, what causes the errors, and what are the associated harms?

Findings In this cohort study of 2428 patient records, a missed or delayed diagnosis took place in 23%, with 17% of these errors causing temporary or permanent harm to patients. The underlying diagnostic process problems with greatest effect sizes associated with diagnostic errors, and which might be an initial focus for safety improvement efforts, were faults in testing and clinical assessment.

Meaning Among hospitalized adults transferred to the ICU or who died in the hospital, diagnostic errors were common, harmful, and had underlying causes, which can be used to design future interventions.

Methods

Study Design

We conducted a retrospective multicenter cohort study of adult patients who died or were transferred to the ICU after the second hospital day. We excluded patients who were transferred to the ICU earlier in their course to eliminate cases due to mistriage from the emergency department rather than inpatient diagnostic errors. This study was reviewed and approved by the University of California San Francisco Institutional Review Board (IRB), with study sites' IRBs relying on that approval under a single IRB mechanism. The informed consent requirement was waived for this study based on the low-risk nature of the study, its retrospective nature, and the fact that many participants would be unable to provide consent (eg, had died).

Sites and Patients

This study was undertaken as a collaboration among 29 academic centers participating in the Hospital Medicine ReEngineering Network (HOMERUN),[19] a national collaborative of academic medical centers including university-based centers, community-based teaching hospitals, and safety-net hospitals. We identified patients (**Figure 1**) by screening administrative data collected from participating sites (Vizient Clinical Data Base; Vizient Inc), yielding an initial cohort of 487 532 patients admitted to participating sites between January 1 and December 31, 2019, and who had a medical diagnosis as defined by the Centers for Medicare & Medicaid Services, of whom 24 591 (5.0%) died or were transferred to the ICU during their hospitalization. The "other" race and ethnicity category included unknown, other, unavailable, and declined, as defined by Vizient. Because some sites were larger than others, we then randomly selected patients within each site's sample to ensure balanced availability of cases for review. Reviewers then screened cases in random order, excluding any patient whose case was identified in error (eg, not a medical diagnosis), whose ICU transfer was for a policy reason (eg, desensitization to a medication), whose admission was for comfort or hospice care only, whose admission followed an out-of-hospital cardiac arrest, or if the medical record was unavailable. This screening yielded 2997 eligible cases, which were reviewed until 100 medical records were adjudicated at each site or the data collection period was completed. After exclusions and reviews were complete, our final cohort included 2428 patients.

Adjudication Methodology

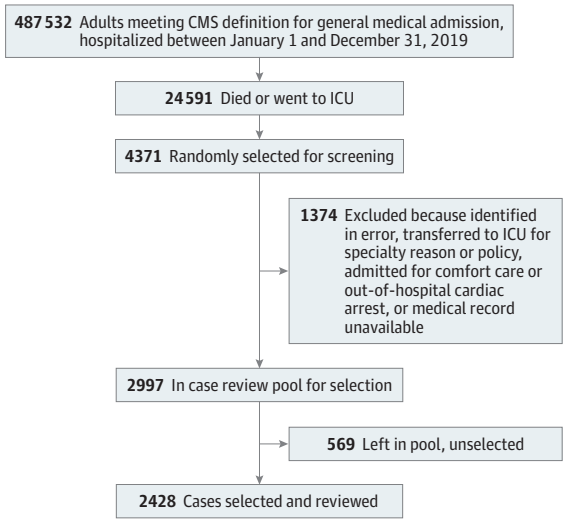
All cases in this study were reviewed by 2 physicians trained in error adjudications, with extensive oversight and quality-checking steps in place.[20] Use of 2 physician reviews is a common approach in patient safety research[15,21] and a method we used in past research examining readmissions[22] and diagnostic errors.[23] Both physician reviewers needed to agree to the entirety of the adjudication assessment for the adjudication to be finalized. If agreement could not be achieved, the pair engaged a third trained reviewer at the site to resolve differences.

The process of 2 (or in some cases, 3) physician reviews by definition produces a case review with complete agreement between trained reviewers, and in that context, we did not measure interrater reliability. Having said this, separate work from our team has demonstrated that adjudications performed by 2 trained physicians not associated with the case, compared with expert overreads, produces results with Cohen κ greater than 0.7 for identifying diagnostic errors.[14,24]

Adjudicator Selection and Training

Reviewers were active clinicians caring for general medical inpatients who were trained to identify diagnostic errors by

Figure 1. Patient Identification, Selection, and Review Processes



Patients could be excluded for more than 1 reason. CMS indicates Centers for Medicare & Medicaid Services; ICU, intensive care unit.

participating in a 2-day live video conference and then reviewing at least 5 standardized cases with expert reviewers. Initial training was followed by independent review of additional standardized cases in blocks of 5 with overreads by members of our research team until we observed 100% agreement.

Adjudication Data Quality Assurance

To ensure consistency across sites, each site presented at least 1 case (including redacted clinical materials and the adjudication forms) to study team members quarterly, where cases received feedback and corrections if needed. Additionally, each site redacted and sent every tenth case for independent expert overread by the research team. Using this process, a minimum of 14 cases per site (more than 500 overall) were confirmed by the research team during the adjudication process.

As a final validity check, the research team directly reexamined a minimum of 10 redacted patient medical records and original case review forms from sites whose error rates were more than 1 SD above or below the group mean error rate (4 sites). These checks confirmed high concordance at all but 1 site; for that site, we retained data from only 23 cases, which were overread and confirmed by 2 additional members of our team.

Determination of Errors and Underlying Causes

Reviewers examined the entire electronic medical record for each hospitalization, with particular focus on the reason for admission and events leading up to ICU transfer or death. In each case, adjudicators strove to correlate documentation regarding diagnostic decision-making to results and timestamps for objective data such as vital signs, laboratory and diagnostic test results, and orders. Every medical record was reviewed for the presence or absence of a diagnostic error and any underlying diagnostic process faults; cases with a diagnostic error were also reviewed for harms attributable to the error.

Diagnostic errors were identified using a slightly modified version of the Safer-Dx algorithm,[20,24,25] a medical record–based approach that identifies cases where a diagnostic error might have taken place. We also reviewed all medical records to gather diagnostic process fault information using the Diagnostic Error Evaluation and Research (DEER)[2,26,27] framework (eTable in Supplement 1), adapted slightly to apply to inpatient-specific scenarios (such as transfers from outside hospitals).[14] DEER diagnostic process faults represent steps that might impede a timely and accurate diagnosis (the outcome of the process). As in other safety problems where gaps in care may or may not result in an adverse event, it is not possible to have a diagnostic error without a process fault, but not every process fault will lead to a diagnostic error; the latter would be the equivalent of a near miss in other areas of patient safety. For example, a specialty consultation may have been ordered late but did not lead to a clinically important delay in making the diagnosis. Cases with errors were reviewed for the harm related to the error using the National Coordinating Council for Medication Error Reporting and Prevention scale, which provides explicit definitions of harm (eg, an error was considered to have led to death if it "contributed to or resulted in the patient's death").[28] We provided case examples and rules for adjudicators regarding degrees of harm associated with diagnostic error.

Outcomes and Predictors

The primary outcome of this study (dependent variable) was the presence or absence of a diagnostic error during the index hospitalization, defined as a missed opportunity to make a correct or timely diagnosis based on the available evidence, regardless of patient harm.[29,30] These would include misdiagnoses in addition to missed or delayed diagnoses. Secondary outcomes included harmful diagnostic errors. The major predictors (independent variables) included DEER taxonomy diagnostic process faults.

Statistical Analysis

The rates of diagnostic error and variation across sites are presented descriptively using means and SDs. We classified comorbidities based on the method of Elixhauser[31] and used *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* codes (in the primary or secondary position) to define inpatient diagnoses commonly associated with diagnostic errors, such as cerebrovascular accidents and sepsis.[2,32] The rates of diagnostic error in patients defined by characteristics obtained from medical record review and administrative data were compared. All analyses, including estimation of univariate proportions and their confidence intervals, involved weighted estimation, with each observation weighted by the inverse of the sampling probability, which was defined as the ratio of cases reviewed in each hospital by the total number of ICU transfers and deaths eligible for review at each hospital during the study period.[33,34]

Multivariable Cox proportional hazard models incorporating the effect of clustering with the time variable set to unity and ties handled with the Breslow method[35] were used to estimate adjusted rate ratios associated with DEER factors. Robust variance estimators were used to construct confidence intervals of parameter estimates. Since the outcomes of error and harmful error were common in these data, the odds ratio is a poor approximate to the prevalence ratio, which is generally considered a more interpretable measure of association in cross-sectional studies. Therefore, we used a modified form of Cox regression to directly estimate the prevalence ratio. When the time to event (at-risk period) is set to an arbitrary value for all observations and the Breslow method of handling ties is used, the hazard ratio estimated by Cox regression is equivalent to the prevalence ratio in cross-sectional studies.[36,37] Covariates for multivariable models were chosen based on substantive knowledge and a priori hypotheses regarding the association between each variable and diagnostic error, as well as the observed association between the variable and the outcome in bivariate analyses.

We calculated adjusted preventable attributable fractions, taking into account the sampling design[38] (ie, the proportion of diagnostic errors that would have been eliminated if that process fault were eliminated) as a way to provide guidance around which features contributed most to diagnostic errors in absolute terms. The χ^2 test or 2-tailed Fisher exact test with simulated *P* values (2000 replicates), with .05 as the level of significance. All analyses were conducted using R Statistical Software, version 4.1.2 (R Core Team 2021), and SAS, version 9.4 (SAS Institute).

Results

Patient Characteristics and Diagnostic Error Rates

Of 2428 patient records at 29 hospitals that underwent review (mean [SD] patient age, 63.9 [17.0] years; 1107 [45.6%] female and 1321 male individuals [54.4%]), 550 patients experienced a diagnostic error, representing a mean (SD) error rate of 23.0% (42.1%), which varied fairly widely across sites. Patient demographics (eg, age, sex, race and ethnicity, and admission source), as well as most administratively coded comorbidities and factors manually identified during adjudication, were statistically similar between groups (**Table 1**). **Table 2** provides illustrative cases of diagnostic errors associated with each of the major categories of DEER diagnostic process faults.

Harms Associated With Diagnostic Errors

Errors were judged to have contributed to temporary harm, permanent harm, or death in 436 patients (17.8%; 95% CI, 15.9%-19.8%) (**Table 3**). The rate of harmful errors was also variable across sites (SD, 38.2%). Among 550 patients with diagnostic error, the error was judged to have contributed to temporary harm, permanent harm, or death in 77.1% (95% CI, 72.3%-81.9%). Among all 1863 patients who died, the diagnostic error was judged to have contributed to death in 121 (6.6%; 95% CI, 5.3%-8.2%); within the group of patients who died and had a diagnostic error, the error contributed to the death in 29.4% (95% CI, 24.0%-35.3%).

Diagnostic Process Faults Associated With Diagnostic Errors

The most prevalent diagnostic process faults in our cohort were problems with assessment (eg, delay in considering diagnosis, failure to recognize complications, or suboptimal prioritizing of potential diagnoses), access and presentation faults (eg, incorrect triage, failure or delay in seeking care), and problems with testing (eg, delay in ordering or performing needed tests, erroneous clinician interpretation of test) (**Figure 2**). In multivariable models adjusting for patient sociodemographic factors, comorbidities, and all process faults, the 2 diagnostic processes most highly associated with diagnostic error were problems with assessment (adjusted relative risk, 2.89; 95% CI, 2.23-3.73) and testing (adjusted relative risk, 2.85; 95% CI, 2.16-3.76), corresponding to adjusted proportion attributable fractions of 21.4% (95% CI, 16.4%-26.4%) and 19.9% (95% CI, 14.7%-25.1%), respectively.

Secondary Analyses

We also conducted prespecified secondary analyses examining the association between DEER features and diagnostic errors limited to sites that did case reviews in at least 90 patients and among sites whose error rate was in the middle 50% of error distribution; results of these analyses were similar to the primary results, suggesting that site-level case numbers or error rates were not associated with the impact of diagnostic process faults. We also undertook analyses examining the association between DEER features and harmful diagnostic errors. In this analysis, confidence intervals widened due to smaller sample size, but a similar ranking of diagnostic process faults with the highest adjusted proportion attributable fractions (assessment, follow-up, and testing) was observed.

Discussion

In this multicenter study of a selected group of medical patients who died in hospital or who were transferred to an ICU, diagnostic errors were common and associated with patient

Table 1. Patient Characteristics (N = 2428)

Characteristic	Patients, No. (%)[a]		*P* value
	Error present (n = 550)	Error absent (n = 1878)	
Information from administrative data			
Sex			
Female	250 (45.5)	857 (45.6)	.94
Male	300 (54.5)	1021 (54.4)	
Age, median (IQR), y	65 (56-76)	66 (55-77)	.40
Race			
Asian	24 (4.4)	94 (5.0)	.17
Black	114 (20.7)	316 (16.8)	
White	367 (66.7)	1290 (68.7)	
Other[b]	45 (8.2)	178 (9.5)	
Ethnicity			
Hispanic	33 (6.0)	99 (5.3)	.39
Non-Hispanic	490 (89.1)	1660 (88.4)	
Unknown	27 (4.9)	119 (6.3)	
Admission status			
Emergency	406 (73.8)	1340 (71.4)	.67
Urgent	124 (22.5)	456 (24.3)	
Elective	14 (2.5)	61 (3.2)	
Trauma center	6 (1.1)	21 (1.1)	
Primary payer			
Commercial	101 (18.4)	377 (20.1)	.02
Medicaid	95 (17.3)	283 (15.1)	
Medicare	337 (61.3)	1103 (58.7)	
Other	17 (3.1)	115 (6.1)	
Death and/or ICU transfer			
Transferred to ICU after 24 h but did not die	144 (26.2)	421 (22.4)	<.001
Inpatient death without ICU transfer	318 (57.8)	1271 (67.7)	
Death after transfer to ICU	88 (16.0)	186 (9.9)	
Comorbidities from administrative data			
Congestive heart failure	221 (40.2)	667 (35.5)	.046
Hypertension, complicated	245 (44.5)	738 (39.3)	.03
Chronic pulmonary disease	140 (25.5)	510 (27.2)	.43
Diabetes, complicated	153 (27.8)	461 (24.5)	.12
Kidney failure	226 (41.1)	677 (36.0)	.03
Liver disease	212 (38.5)	676 (36.0)	.28
Metastatic cancer	63 (11.5)	273 (14.5)	.07
Coagulopathy	263 (47.8)	804 (42.8)	.04
Obesity	96 (17.5)	259 (13.8)	.03
Weight loss	228 (41.5)	680 (36.2)	.02
Fluid and electrolyte disorders	489 (88.9)	1566 (83.4)	.002
Alcohol use disorder	100 (18.2)	268 (14.3)	.02
Substance use disorder	45 (8.2)	120 (6.4)	.14
Diagnostic error-prone principal diagnoses[39]			
Sepsis	272 (49.5)	946 (50.4)	.70
Stroke	40 (7.3)	99 (5.3)	.08
Myocardial infarction	69 (12.5)	224 (11.9)	.70

(continued)

Table 1. Patient Characteristics (N = 2428) (continued)

Characteristic	Patients, No. (%)[a]		*P* value
	Error present (n = 550)	Error absent (n = 1878)	
Information from medical record review			
Patient or caregiver preferences for care affected the diagnostic process	53 (10.8)	256 (16.2)	.003
Prior to admission, patient had a primary care physician or other regular source of outpatient care	419 (77.2)	1403 (75.7)	.48
Housing instability or unhoused	29 (5.3)	47 (2.5)	.001
English is the patient's primary language	477 (86.7)	1637 (87.3)	.72
Any other barriers to communication	173 (31.5)	727 (38.7)	.002
Altered mental status on presentation	223 (40.6)	890 (47.7)	.003

Abbreviation: ICU, intensive care unit.

[a] Counts and percentages are unweighted.

[b] Other included unknown, other, unavailable, and declined.

harm. Problems related to testing, such as choosing the correct test, ordering the test in a timely fashion, or correctly interpreting results, and problems with assessment, such as recognizing complications or revisiting a differential diagnosis, appear to be the most important targets for safety improvement programs.

Estimates of diagnostic error rates vary widely. The Harvard Medical Practice Study[9,10] provided preliminary outlines of the prevalence of diagnostic errors but focused on procedural complications and medication errors as contributors to adverse events; follow-up studies also did not screen for diagnostic errors specifically and likely underestimated their prevalence.[15,21] Meta-analyses of studies reporting diagnostic errors determined via a range of methods suggested rates of 10% or lower,[18] while autopsy studies have described rates between 5% and 25%.[16,17,40] Finally, recent work calculated a national prevalence closer to 20% using a combination of administrative data and literature-based rates.[39] Our results fall in the upper end of the range defined in previous studies and provide additional insights in several important ways. First, we used a standard approach to identifying our patient cohort, focusing the scope of diagnostic review on patients who experienced similar clinical events. Second, we examined medical records in detail and applied a rigorous adjudication process, which permitted us to directly measure diagnostic errors in a reliable and valid way, rather than inferring the presence of an error based on combinations of events. Third, although deaths and ICU transfers are statistically infrequent and likely represent a seriously ill patient population, the importance of these events in patient safety efforts is paramount, making our results immediately useful to hospitals focused on addressing these events. Finally, although we observed wide variations in diagnostic error rates across sites, similarly wide ranges of errors across sites have also been seen in other studies of diagnostic errors,[41] an observation that might influence strategies to prevent or mitigate errors.

While aspects of the diagnostic process, such as weighing alternate diagnoses or conferring with colleagues, may take place outside of the electronic health record, documentation remains a key means of communication between clinicians, patients, and families. Directly capturing communication or cognitive processes (via observation, surveys, or interviews) poses its own challenges due to recall biases, the Hawthorne effect, or "second-victim" harms.[42] Testing problems and gaps in clinical assessment may be fruitful targets for future interventions seeking to reduce missed or delayed diagnoses. Our data are not of sufficient granularity to discern specific tests or testing scenarios but may help narrow future interventions. Solutions to testing problems may rely in large degree on informatics tools such as alerts or predictive models. In contrast, clinical assessment gaps may also require evaluation of physician workload, as well as coaching, debiasing, and cognitive interventions,[43] such as diagnostic timeouts[44] or systems that prompt clinicians to consider alternate diagnoses.[45] Emergence of artificial intelligence and large language models hold promise through their ability to gather and synthesize complex data necessary to make an accurate and timely diagnosis. Our methods can assist in the development of advanced models by providing criterion-standard error reviews needed to build models and assess the effect of interventions, not to mention an approach to continuously monitoring model output to avoid inaccuracies and biases.

Limitations

Our study has several limitations. Our results do not represent the prevalence and severity of diagnostic errors across all hospitalized patients, as this was a select sample of patients who experienced clinical deterioration. Moreover, these results may not be generalizable to all US hospitals, given the selection of mostly academic medical centers for this study. Our data are subject to documentation and detection biases. To overcome documentation biases, we encouraged medical record reviewers to use all available documentation in the medical record (eg, notes, test results, orders) and to use reasonable judgment to interpret patterns seen as indicative of the diagnostic process. To further address detection biases, all reviewers underwent extensive training at study outset, leveraging methods that have been shown to produce high interrater reliability.[14,24] To increase validity across sites, cases were overread and reviewed by

Table 2. Case Vignettes of Diagnostic Errors With Examples of Process Faults[a]

DEER process fault dimensions	Case vignette, associated DEER dimension, and other faults
Access/presentation	A patient with severe aortic stenosis was transferred from an outside hospital to the surgical service for evaluation of odontoid fracture with worsening posterior displacement. The patient was tachycardic prior to arrival and both hypotensive and tachycardic immediately on arrival. Progressive shock prompted medicine ICU transfer 36 h later, where the patient was diagnosed with low-flow, low-gradient aortic stenosis ascribed to sepsis or cardiogenic causes. The patient's condition became increasingly difficult to manage with pressors, and the patient ultimately died after being transitioned to comfort care. Dimension related: access faults—incorrect triage to surgical service rather than critical care or medical service Other faults: failure or delay in performing needed test(s)—earlier full evaluation for shock
History taking	A patient was admitted with fecal impaction and acute kidney injury thought to be due to hypovolemia, initially treated with intravenous hydration and magnesium citrate. On the second day, the patient was transferred to the ICU for hypotension and bradycardia in the setting of a serum magnesium level of 10.2 mg/dL. It was later learned that the patient had been taking frequent magnesium citrate at home. Dimension related: history taking process fault—failure or delay in providing or eliciting a critical piece of history data Other faults: failure or delay in ordering needed test(s)—serum magnesium level
Physical examination	A patient was transferred from an outside hospital with bilateral lower extremity weakness progressing over 3 d. Physical examination documented decreased reflexes and lower extremity weakness. The clinical picture initially ascribed weakness to anxiety and somatization rather than organic cause and did not explain loss of reflexes. A lumbar puncture was performed 1 d later, but the patient had a cardiac arrest that night and ultimately died; cerebrospinal fluid results confirmed Guillain-Barre syndrome. Dimension related: physical examination process faults—inaccurate or misinterpreted physical examination finding and suboptimal weighing of a physical examination finding Other faults: failure or delay in ordering monitoring
Testing	A patient receiving long-term anticoagulation therapy was admitted with a left psoas hematoma following bone marrow biopsy. Anticoagulation therapy was restarted on hospital day 5, after which tachycardia and increased back and left lower extremity pain recurred. A CT angiogram was not ordered until the next morning and was not performed for an additional 9 h. When obtained, the CT revealed active extravasation prompting IR intervention. Dimension related: testing fault—failure or delay in ordering CT angiogram Other faults: assessment—failure or delay in recognizing complications
Patient follow-up and monitoring	A patient was admitted with group B streptococcal septicemia due to soft tissue infection of the foot. Initial care plans focused on concern for meningitis and lacked a plan for monitoring of the foot infection. The patient was transferred to the ICU for worsening hemodynamics, after which surgical debridement of the foot was performed. Dimension related: follow-up and monitoring fault—failure or delay in recognizing or acting on urgent condition or complications of soft tissue infection Other faults: failure or delay in performing needed test for deep tissue infection, suboptimal weighing or prioritizing (eg, too much weight given to lower-probability diagnoses)
Consultation and referral	A patient was transferred to the medicine service for hypoxemia and abdominal pain after a diagnostic hysteroscopy by gynecology service. Abdominal CT showed free air, thought by the gynecology consult to be due to benign uterine perforation. The patient clinically deteriorated with lactic acidosis, hypotension requiring vasopressors, and worsening hypoxemia, prompting engagement of the general surgical team, who brought the patient emergently to the operating room and diagnosed a small bowel perforation. Dimension related: consultation and referral fault—problems with original consulting service decision-making, delay in involvement of general surgery Other faults: suboptimal weighing or prioritizing (eg, too little weight given to higher-probability or higher-risk diagnosis)
Teamwork	A patient with peripheral arterial disease was admitted with dry gangrene of the foot. They were transferred to the ICU for hypercarbic and hypoxemic respiratory failure on the second hospital day after an evening where they were noted by team members to be more somnolent and confused. The change in mental status noted by nursing and respiratory therapy was not communicated to the primary team until immediately before transfer to the ICU. Dimension related: teamwork fault—failure or delay in communication of information within the patient's care team Other faults: missed physiologic monitoring finding (eg, persistent hypoxia, oxygen requirement)
Patient communication and experience	A patient with Parkinson disease presented with aspiration pneumonia and a pleural effusion that appeared exudative in nature. They were treated with antibiotics with resolution of symptoms and discharged. Pleural fluid cytology showed malignant cells, not documented in the discharge summary or communicated to the patient's family. Dimension related: communication fault—failure or delay in communicating laboratory or test results, assessment, or consultant findings to the patient/caregiver Other faults: failure or delay in acting on or following up on test result
Assessment	A patient with recently diagnosed hepatitis C–associated vasculitis was admitted with hematochezia thought due to NSAID use for pain along with moderate anion gap acidosis. The patient was transferred to the ICU with continued acidemia, encephalopathy, and hypoxemic respiratory failure initially ascribed to noncardiogenic pulmonary edema or pneumonia. Salicylate toxicity, which was not considered at admission, was diagnosed after ICU transfer by serum salicylate levels. Dimension related: assessment fault—failure or delay in considering the diagnosis Other faults: failure or delay in ordering needed test(s)

Abbreviations: CT, computed tomography; DEER, Diagnostic Error Evaluation and Research framework; ICU, intensive care unit; IR, interventional radiology; NSAID, nonsteroidal anti-inflammatory drug.

[a] Process faults listed are not presented in terms of relative importance or in the order in which they might have taken place, but represent those considered present and related to the diagnostic error.

Table 3. Severity of Harms Associated With Diagnostic Errors (n = 550 Errors)

Error type	No.	Prevalence,[a] % (95% CI)[b,c]
Error did not reach the patient	15	1.9 (1.2-3.2)
Error reached the patient but did not cause harm	64	12.8 (9.5-17.1)
Error required monitoring to confirm that it resulted in no harm	35	8.1 (5.3-12.2)
Error may have contributed to or resulted in temporary harm and required intervention	91	14.2 (11.1-18.0)
Error may have contributed to or resulted in temporary harm and required initial or prolonged hospitalization	116	21.0 (16.9-25.9)
Error may have contributed to or resulted in permanent harm	70	11.7 (8.8-15.4)
Error required intervention necessary to sustain life	31	6.9 (4.4-10.7)
Error may have contributed to or resulted in the patient's death	128	23.3 (19.1-28.1)

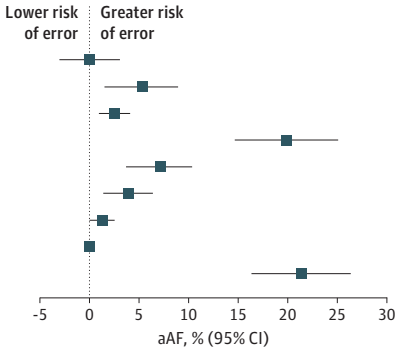
[a] Prevalences weighted by the inverse of the sampling fraction, or the number of cases reviewed at each site divided by the total number of deaths and intensive care unit transfers at each site.

[b] Confidence intervals were not adjusted for multiplicity and should not be used in place of hypothesis testing.

[c] Percentages may not total 100 because of rounding.

Figure 2. DEER Process Fault Dimensions: Prevalence, Adjusted Associations With Diagnostic Errors, and Adjusted Attributable Fractions (aAFs) (N = 2428)

Diagnostic process dimensions	Patients, No. (%)		aRR (95% CI)	aAF, % (95% CI)
	No error	Error		
Access/presentation	232 (14.1)	36 (8.8)	0.79 (0.60 to 1.05)	0.0 (−3.1 to 3.0)
History taking	80 (5.0)	54 (12.9)	1.71 (1.23 to 2.36)	5.2 (1.6 to 8.9)
Physical examination	19 (0.9)	45 (7.8)	1.86 (1.39 to 2.49)	2.5 (0.9 to 4.1)
Testing	60 (4.4)	158 (34.1)	2.85 (2.16 to 3.76)	19.9 (14.7 to 25.1)
Patient follow-up and monitoring	47 (3.0)	81 (15.4)	1.94 (1.45 to 2.60)	7.0 (3.7 to 10.4)
Obtaining referrals	42 (1.9)	58 (10.0)	1.54 (1.13 to 2.09)	3.8 (1.3 to 6.3)
Teamwork	4 (0.1)	11 (2.3)	2.89 (1.62 to 5.16)	1.2 (−0.1 to 2.5)
Communication with patient/caregiver	5 (0.3)	2 (0.3)	1.47 (0.66 to 3.32)	0.0 (−0.2 to 0.2)
Assessment	87 (3.7)	220 (37.1)	2.89 (2.23 to 3.73)	21.4 (16.4 to 26.4)



Multivariable models included adjustment for sex, race, ethnicity, admission source, admission status, insurance (primary payer); the following comorbidities: congestive heart failure, complicated hypertension, kidney failure, chronic pulmonary disease, complicated diabetes, fluid and electrolyte disorders, liver disease, metastatic cancer, obesity, alcohol use disorder, substance use disorder; a primary diagnosis of sepsis, stroke, or myocardial infarction; whether patient preferences affected the diagnostic process; and whether the patient had a primary care physician, housing challenges, communication challenges, or altered mental status. Adjusted rate ratios (aRRs) were estimated using Cox proportional hazard models, with the time variable set to unity for all individuals, using the Breslow method for ties. Adjusted attributable fractions were computed using logistic regression models. DEER indicates the Diagnostic Error Evaluation and Research framework.

members of the core research team, and data were cross-checked extensively. Our data cannot distinguish what type of cognitive process was associated with a diagnostic error (for example, anchoring on a diagnosis to the exclusion of others). The medical record also likely underdetects communication gaps or issues with team dynamics, thus explaining the low prevalence of these issues in our study. For similar reasons, we cannot assess whether patients experienced different sorts of harm (such as emotional or financial harms) related to diagnostic errors. It is possible that local reviewers' adjudications were shaped by local norms and professional standards (eg, expectations for consultation timeliness), or that assessments of the likelihood of an error or its attendant harms were influenced by the fact that all patients had experienced an ICU transfer or death. We addressed both problems via training and during intersite overread of cases. We cannot disentangle the association between clinical assessment and other faults. For example, it is possible that testing process faults might lead to problems in clinical assessment, or the reverse. Our study was not able to directly measure external pressures on teams or clinicians that might affect cognitive processes, such as hospital census or physician workload.

Conclusions

In this cohort study of hospitalized patients who died or were transferred to the ICU, diagnostic errors were common, harmful, and associated with factors that can become potential opportunities for interventions. Results from our study provide impetus for rapid exploration and testing of interventions seen to reduce diagnostic errors and harms associated with ICU transfers and deaths by targeting gaps in test selection and interpretation and physicians' ability to debias and rethink diagnoses as high-priority areas.

ARTICLE INFORMATION

Accepted for Publication: November 7, 2023.

Published Online: January 8, 2024. doi:10.1001/jamainternmed.2023.7347

Author Affiliations: Division of Hospital Medicine, Department of Medicine, University of California San Francisco (Auerbach, Lee, Hubbard); Division of Hospital Medicine, Zuckerberg San Francisco General Hospital, San Francisco, California (Ranji); Department of Medicine, University of Colorado School of Medicine, Denver (Raffel); Department of Radiation Oncology, University of California San Francisco (Valdes); Division of Geriatrics, Department of Medicine, University of California San Francisco (Boscardin); Hospital Medicine Unit, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts (Dalal, Schnipper); Vizient Inc, Irving, Texas (Harris, Flynn).

Author Contributions: Drs Auerbach and Hubbard 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.
Concept and design: Auerbach, Ranji, Raffel, Dalal, Harris, Flynn, Schnipper.
Acquisition, analysis, or interpretation of data: Auerbach, Lee, Hubbard, Ranji, Valdes, Boscardin, Dalal, Harris, Schnipper.
Drafting of the manuscript: Auerbach, Hubbard, Ranji, Valdes, Dalal, Schnipper.
Critical review of the manuscript for important intellectual content: Auerbach, Lee, Hubbard, Ranji, Raffel, Boscardin, Dalal, Harris, Flynn, Schnipper.
Statistical analysis: Auerbach, Hubbard, Valdes, Boscardin, Dalal, Harris.
Obtained funding: Auerbach, Schnipper.
Administrative, technical, or material support: Auerbach, Lee, Ranji, Raffel, Dalal, Harris, Flynn.
Supervision: Auerbach, Ranji, Dalal, Schnipper.

Conflict of Interest Disclosures: Dr Auerbach reported being a founder of Kuretic Health outside the submitted work. Ms Lee reported grants from Agency for Healthcare Research and Quality (AHRQ) during the conduct of the study. Dr Hubbard reported grants from AHRQ during the conduct of the study. Dr Ranji reported grants from AHRQ during the conduct of the study. Dr Dalal reported grants from AHRQ and CRICO during the conduct of the study; equity from I-PASS Institute outside the submitted work; in addition, Dr Dalal had a patent for Real-Time Diagnostic Error Prediction Algorithm pending, a patent for Diagnostic Time-Out pending, and a patent for Patient Diagnostic Questionnaire pending. Ms Flynn reported grants from AHRQ during the conduct of the study. Dr Schnipper reported grants from AHRQ during the conduct of the study. No other disclosures were reported.

Funding/Support: This study was supported by AHRQ grant R01HS027369.

Role of the Funder/Sponsor: AHRQ had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: The members of the UPSIDE Research Group are listed in Supplement 2.

Data Sharing Statement: See Supplement 3.

Additional Contributions: In memory of David Meyers, MD: We would like to acknowledge Dr David Meyers, for his support of our work and our research network since this project's inception.

REFERENCES

1. National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*. National Academies Press; 2015. doi:10.17226/21794

2. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med*. 2009;169(20):1881-1887. doi:10.1001/archinternmed.2009.333

3. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ*. 2010;44(1):94-100. doi:10.1111/j.1365-2923.2009.03507.x

4. Scarpello J. Diagnostic error: the Achilles' heel of patient safety? *Clin Med (Lond)*. 2011;11(4):310-311. doi:10.7861/clinmedicine.11-4-310

5. Ely JW, Kaldjian LC, D'Alessandro DM. Diagnostic errors in primary care: lessons learned. *J Am Board Fam Med*. 2012;25(1):87-97. doi:10.3122/jabfm.2012.01.110174

6. Singh H, Graber ML, Kissam SM, et al. System-related interventions to reduce diagnostic errors: a narrative review. *BMJ Qual Saf*. 2012;21(2):160-170. doi:10.1136/bmjqs-2011-000150

7. Groszkruger D. Diagnostic error: untapped potential for improving patient safety? *J Healthc Risk Manag*. 2014;34(1):38-43. doi:10.1002/jhrm.21149

8. McCarthy M. Diagnostic error remains a pervasive, underappreciated problem, US report says. *BMJ*. 2015;351:h5064. doi:10.1136/bmj.h5064

9. Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II. *N Engl J Med*. 1991;324(6):377-384. doi:10.1056/NEJM199102073240605

10. Brennan TA, Leape LL, Laird NM, 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(6):370-376. doi:10.1056/NEJM199102073240604

11. Saber Tehrani AS, Lee H, Mathews SC, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank. *BMJ Qual Saf*. 2013;22(8):672-680. doi:10.1136/bmjqs-2012-001550

12. Bhise V, Sittig DF, Vaghani V, Wei L, Baldwin J, Singh H. An electronic trigger based on care escalation to identify preventable adverse events in hospitalised patients. *BMJ Qual Saf*. 2018;27(3):241-246. doi:10.1136/bmjqs-2017-006975

13. Hanskamp-Sebregts M, Zegers M, Vincent C, van Gurp PJ, de Vet HC, Wollersheim H. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open*. 2016;6(8):e011078. doi:10.1136/bmjopen-2016-011078

14. Griffin JA, Carr K, Bersani K, et al. Analyzing diagnostic errors in the acute setting: a process-driven approach. *Diagnosis (Berl)*. 2021;9(1):77-88. doi:10.1515/dx-2021-0033

15. Bates DW, Levine DM, Salmasian H, et al. The safety of inpatient health care. *N Engl J Med*. 2023;388(2):142-153. doi:10.1056/NEJMsa2206117

16. Schwanda-Burger S, Moch H, Muntwyler J, Salomon F. Diagnostic errors in the new millennium: a follow-up autopsy study. *Mod Pathol*. 2012;25(6):777-783. doi:10.1038/modpathol.2011.199

17. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA*. 2003;289(21):2849-2856. doi:10.1001/jama.289.21.2849

18. Gunderson CG, Bilan VP, Holleck JL, et al. Prevalence of harmful diagnostic errors in hospitalised adults: a systematic review and meta-analysis. *BMJ Qual Saf*. 2020;29(12):1008-1018. doi:10.1136/bmjqs-2019-010822

19. Auerbach AD, Patel MS, Metlay JP, et al. The Hospital Medicine Reengineering Network (HOMERuN): a learning organization focused on improving hospital care. *Acad Med*. 2014;89(3):415-420. doi:10.1097/ACM.0000000000000139

20. Dalal AK, Schnipper JL, Raffel K, Ranji S, Lee T, Auerbach A. Identifying and classifying diagnostic errors in acute care across hospitals: early lessons from the Utility of Predictive Systems in Diagnostic Errors (UPSIDE) study. *J Hosp Med*. Published online May 21, 2023. doi:10.1002/jhm.13136

21. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med*. 2010;363(22):2124-2134. doi:10.1056/NEJMsa1004404

22. Auerbach AD, Kripalani S, Vasilevskis EE, et al. Preventability and causes of readmissions in a national cohort of general medicine patients. *JAMA Intern Med*. 2016;176(4):484-493. doi:10.1001/jamainternmed.2015.7863

23. Auerbach AD, Astik GJ, O'Leary KJ, et al. Prevalence and causes of diagnostic errors in hospitalized patients under investigation for COVID-19. *J Gen Intern Med*. 2023;38(8):1902-1910. doi:10.1007/s11606-023-08176-6

24. Malik MA, Motta-Calderon D, Piniella N, et al. A structured approach to EHR surveillance of diagnostic error in acute care: an exploratory analysis of two institutionally-defined case cohorts. *Diagnosis (Berl)*. 2022;9(4):446-457. doi:10.1515/dx-2022-0032

25. Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. *BMJ Qual Saf*. 2015;24(2):103-110. doi:10.1136/bmjqs-2014-003675

26. Schiff GD. Diagnosis and diagnostic errors: time for a new paradigm. *BMJ Qual Saf*. 2014;23(1):1-3. doi:10.1136/bmjqs-2013-002426

27. 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, et al, eds. *Advances in Patient Safety: From Research to Implementation*. Vol 2: Concepts and Methodology. Agency for Healthcare Research and Quality; 2005.

28. National Coordinating Council for Medication Error Reporting and Prevention. NCC MERP index for categorizing medication errors (revised 2001).

Accessed October 31, 2023. https://www.nccmerp.org/types-medication-errors

29. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. *BMJ Qual Saf*. 2017;26(6):484-494. doi:10.1136/bmjqs-2016-005401

30. Singh H. Editorial: Helping health care organizations to define diagnostic errors as missed opportunities in diagnosis. *Jt Comm J Qual Patient Saf*. 2014;40(3):99-101. doi:10.1016/S1553-7250(14)40012-6

31. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27. doi:10.1097/00005650-199801000-00004

32. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in malpractice claims: the "big three"—vascular events, infections, and cancers. *Diagnosis (Berl)*. 2019;6(3):227-240. doi:10.1515/dx-2019-0019

33. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26:404-413. doi:10.1093/biomet/26.4.404

34. Sison CP, Glaz J. Simultaneous confidence intervals and sample size determination for multinomial proportions. *J Am Stat Assoc*. 1995;90(429):366-369. doi:10.1080/01621459.1995.10476521

35. Breslow N. Covariance analysis of censored survival data. *Biometrics*. 1974;30(1):89-99. doi:10.2307/2529620

36. Xie W, Zheng F. Robust Cox regression as an alternative method to estimate adjusted relative risk in prospective studies with common outcomes. *Int J Stat Med Res*. 2016;5(4):231-239. doi:10.6000/1929-6029.2016.05.04.1

37. Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol*. 2003;3(1):21. doi:10.1186/1471-2288-3-21

38. Heeringa SG, Berglund PA, West BT, Mellipilán ER, Portier K. Attributable fraction estimation from complex sample survey data. *Ann Epidemiol*. 2015;25(3):174-178. doi:10.1016/j.annepidem.2014.11.007

39. Newman-Toker DE, Nassery N, Schaffer AC, et al. Burden of serious harms from diagnostic error in the USA. *BMJ Qual Saf*. Published online July 17, 2023. doi:10.1136/bmjqs-2021-014130

40. Bergl PA, Zhou Y. Diagnostic error in the critically ill: a hidden epidemic? *Crit Care Clin*. 2022;38(1):11-25. doi:10.1016/j.ccc.2021.09.005

41. Newman-Toker DE, Peterson SM, Badihian S, et al. *Diagnostic Errors in the Emergency Department: A Systematic Review*. AHRQ Comparative Effectiveness Review No. 258 Agency for Healthcare Research and Quality; 2022. doi:10.23970/AHRQEPCCER258

42. Wu AW. Medical error: the second victim. the doctor who makes the mistake needs help too. *BMJ*. 2000;320(7237):726-727. doi:10.1136/bmj.320.7237.726

43. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. *BMJ Qual Saf*. 2013;22(suppl 2):ii65-ii72. doi:10.1136/bmjqs-2012-001713

44. Yale S, Cohen S, Bordini BJ. Diagnostic time-outs to improve diagnosis. *Crit Care Clin*. 2022;38(2):185-194. doi:10.1016/j.ccc.2021.11.008

45. Ramnarayan P, Cronje N, Brown R, et al. Validation of a diagnostic reminder system in emergency medicine: a multi-centre study. *Emerg Med J*. 2007;24(9):619-624. doi:10.1136/emj.2006.044107