Adverse diagnostic events in hospitalised patients: a single-centre, retrospective cohort study

Anuj K Dalal ,[1,2,3] Savanna Plombon,[1,3] Kaitlyn Konieczny,[1] Daniel Motta-Calderon,[1,4] Maria Malik,[1,5] Alison Garber,[1,6] Alyssa Lam,[1] Nicholas Piniella,[1] Marie Leeson,[1] Pamela Garabedian,[1,3] Abhishek Goyal ,[1,3] Stephanie Roulier,[1,3] Cathy Yoon,[1] Julie M Fiskio,[3] Kumiko O Schnock ,[1,2] Ronen Rozenblum,[1,2] Jacqueline Griffin,[7] Jeffrey L Schnipper ,[1,2,3] Stuart Lipsitz,[1,2] David W Bates ,[1,2,3] Patient Safety Learning Laboratory Adjudicator Group

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjqs-2024-017183).

For numbered affiliations see end of article.

Correspondence to
Dr Anuj K Dalal;
adalal1@bwh.harvard.edu

Received 30 January 2024
Accepted 12 August 2024

To cite: Dalal AK, Plombon S, Konieczny K, *et al*. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2024-017183

ABSTRACT

Background Adverse event surveillance approaches underestimate the prevalence of harmful diagnostic errors (DEs) related to hospital care.

Methods We conducted a single-centre, retrospective cohort study of a stratified sample of patients hospitalised on general medicine using four criteria: transfer to intensive care unit (ICU), death within 90 days, complex clinical events, and none of the aforementioned high-risk criteria. Cases in higher-risk subgroups were over-sampled in predefined percentages. Each case was reviewed by two adjudicators trained to judge the likelihood of DE using the Safer Dx instrument; characterise harm, preventability and severity; and identify associated process failures using the Diagnostic Error Evaluation and Research Taxonomy modified for acute care. Cases with discrepancies or uncertainty about DE or impact were reviewed by an expert panel. We used descriptive statistics to report population estimates of harmful, preventable and severely harmful DEs by demographic variables based on the weighted sample, and characteristics of harmful DEs. Multivariable models were used to adjust association of process failures with harmful DEs.

Results Of 9147 eligible cases, 675 were randomly sampled within each subgroup: 100% of ICU transfers, 38.5% of deaths within 90 days, 7% of cases with complex clinical events and 2.4% of cases without high-risk criteria. Based on the weighted sample, the population estimates of harmful, preventable and severely harmful DEs were 7.2% (95% CI 4.66 to 9.80), 6.1% (95% CI 3.79 to 8.50) and 1.1% (95% CI 0.55 to 1.68), respectively. Harmful DEs were frequently characterised as delays (61.9%). Severely harmful DEs were frequent in high-risk cases (55.1%). In multivariable models, process failures in assessment, diagnostic testing, subspecialty consultation, patient experience, and history were significantly associated with harmful DEs.

Conclusions We estimate that a harmful DE occurred in 1 of every 14 patients hospitalised on general medicine, the majority of which were preventable. Our findings underscore the need for novel approaches for adverse DE surveillance.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Adverse diagnostic events (DEs) are underrecognised in hospitalised patients using current surveillance approaches.

WHAT THIS STUDY ADDS

⇒ Based on a weighted random sample and a structured electronic health record-based case review process using validated instruments for assessing the likelihood of harmful DE, it was estimated that about 7% of hospitalised patients who received general medical care experienced an adverse DE within 90 days of admission.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Patient safety research and institutional quality and safety programmes should consider structured case reviews and novel approaches for improving detection of adverse DEs for hospitalised patients.

INTRODUCTION

Diagnostic errors (DEs), defined by the National Academy of Medicine (NAM) as "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] are difficult to detect and characterise. Consequently, their spectrum of harm is variable and underrecognised in patient safety research.[1] In hospitalised patients, failures in diagnostic processes such

as history-taking, testing or assessment often lead to adverse events (AEs) with severe and immediate impact, such as care escalation or death.[2] They can also lead to less severe or more delayed impact.

Systematic reviews of retrospective studies estimate that adverse diagnostic events occur in 0.7% of inpatients, but these are largely based on cohorts with severe outcomes only[3]; thus, these are likely underestimates.[4] In a recent study which relied on the Institute for Healthcare Improvement (IHI) global trigger tool,[5] Bates *et al* estimated that one in four hospitalisations were associated with an AE. Of nearly 1000 AEs detected across 11 hospitals, just 10 DEs (1%) were identified as culprits.[6] These results suggest that current trigger tools alone are likely insensitive for detecting harmful DEs, including cases with less severe outcomes.[5–7] Indeed, studies using the validated Safer Dx instrument have observed higher percentages of harmful DEs for hospitalised patients who were critically ill or readmitted.[8–10]

Using the Safer Dx framework,[11] we developed and validated a structured case review process to train clinicians to use the electronic health record (EHR) to evaluate the diagnostic process during the hospital encounter, assess the likelihood of DE, and characterise the impact and severity of harm.[2 12 13] This process was validated in two cohorts of patients who expired in the hospital and detected harmful DEs in cases of death judged to be preventable and non-preventable by our institutional mortality review process.[12] This process was further validated in a retrospective, multicentre study and found that harmful DEs occurred in 18% of hospitalised patients who died or transferred to the intensive care unit (ICU).[13 14]

In this study, we evaluated a weighted sample of patients hospitalised on general medicine identified by querying the EHR using clinical screening criteria for inpatient DEs to estimate the prevalence of harmful DEs in this population.[11 15 16] Secondarily, we sought to characterise the types of process failures associated with harmful DEs to enhance surveillance approaches and develop preventative interventions.[17]

METHODS

Study design, setting and eligibility

We conducted a retrospective cohort study, approved by the Mass General Brigham Humans Subjects Committee. Adult patients (>18 years old) hospitalised on the general medicine service for at least 24 hours at a large academic medical centre in Boston, MA, USA were identified by querying our enterprise data warehouse (EDW), which received nightly updates from our EHR (Epic Systems, Inc.) between July 2019 and September 2021. Patients who were admitted directly to hospice care (comfort measures only) were not identified by our queries.

Cases were excluded if the patient had a length of stay (LOS) greater than 21 days as described in our validation study.[12] Because of our focus on the diagnostic process related to general medical care, we excluded cases in which the patient was admitted to a general medicine team but received subspecialty care under supervision of a subspecialty attending. For example, we excluded patients typically admitted to a dedicated oncology team for chemotherapy or management of an oncologic complication but overflowed to a general medicine team and received care under the direction of an oncology service attending.

Cases were also excluded if the patient was admitted between April 2020 and December 2020, the time period when our institution experienced major disruptions in hospital operations due to the COVID-19 pandemic. These disruptions included major changes to admitting teams, services and units requiring changes to our EHR starting in early April 2020, precluding our ability to accurately query our EDW and stratify eligible cases using the criteria defined below. For example, because ICU care was moved to a different location in our hospital, cases of patients who transferred to the ICU were misclassified by our EDW queries as many ICU units were repurposed as general medical-surgical units in the EHR. Other factors considered included changes in patient population (eg, demographics changes, declines in admissions[18]) and abrupt changes in team structures and composition (ie, clinicians caring for patients on a dedicated COVID-19 team or outside their expertise[19]) during the initial pandemic waves prior to the availability of vaccines.[20 21] By December 2020, the infrastructure for caring for COVID-19 patients was established, hospital operations had started to normalise (COVID-19 patients were admitted to dedicated teams; elective surgeries resumed; declining need to repurpose clinical staff), vaccines were aggressively being distributed, and our EDW queries reflected correct team, service and unit assignments.

Stratified case sampling approach

Given the limited availability of data regarding harmful DE rates in hospitalised patients, we used a weighted case sampling approach to examine specific subgroups, drawing from the literature, emerging research and expert consensus regarding the predictive value of clinical screening criteria reported by Shenvi and El-Kareh.[10 12 22 23] Cases that met eligibility criteria (figure 1) during each month of the study period were categorised into one of four subgroups. Cases were considered high-risk if the patient (1) transferred to ICU 24 hours or more after admission independent of death, (2) expired within 90 days of admission (either during the hospital encounter or after discharge identified using our EHR's death date) or (3) had complex clinical events but did not transfer to the ICU or expire within 90 days of admission. Cases with complex clinical events had one of a group of triggers including clinical deterioration signs (eg, new or worsening

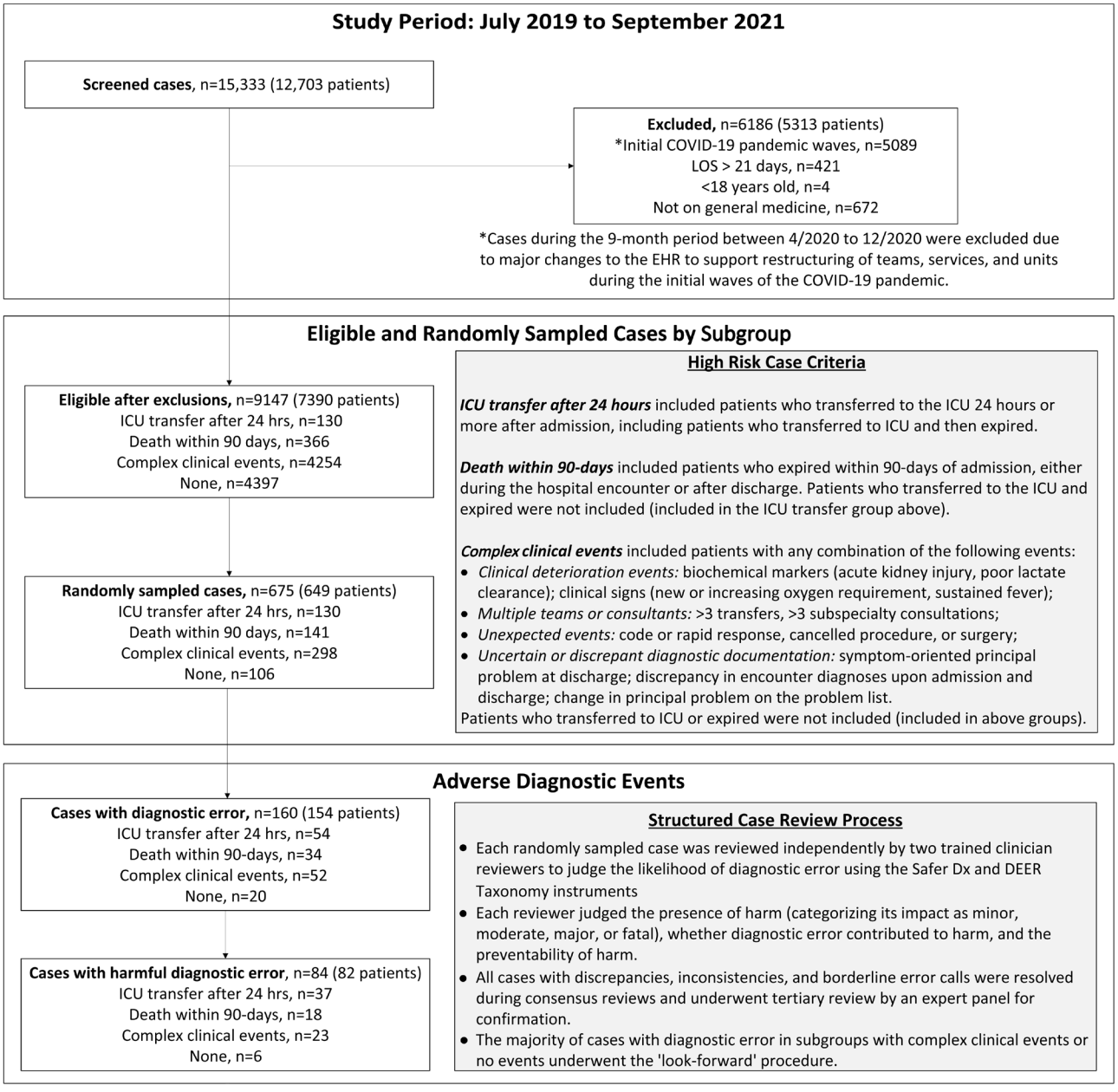

Figure 1 Eligibility and sampling of cases to detect adverse diagnostic events. DEER, Diagnostic Error Evaluation and Research; EHR, electronic health record; ICU, intensive care unit; LOS, length of stay.

oxygen requirement, acute kidney injury, etc.), multiple providers or consultants, unexpected events (eg, rapid response, surgery) or uncertain or discrepant diagnostic documentation, all which were more frequently observed in cases with DE in our validation study.[12 23] Cases without the aforementioned high-risk criteria were considered low-risk (though these could still have an unexpected post-discharge event such as an emergency department visit or readmission). All classifications were made 90 days after admission to ensure appropriate categorisation.

A priori we hypothesised that each subgroup would have a different percentage of harmful DEs. Informed by subanalyses of local data from our validation study and autopsy studies, we expected these percentages to be 30% in the ICU transfer and 20% in the death within 90 days subgroups, respectively.[12 22] For subgroups in which patients did not transfer to the ICU or expire, we assumed the expected percentage would be approximately 5%, largely based on a study of early readmissions by Raffel *et al*.[10] We assumed a higher range (5–10%) in patients with complex clinical events,[23] and a lower range (2–5%) in patients with no events.

Our primary goal was to oversample higher-risk subgroups to gain more information about the probability of harmful DEs while creating a high-yield process for our adjudicators. Our secondary goal was to have sufficient sample size within each subgroup while oversampling the complex clinical events

subgroup (on an absolute basis) relative to the others as it comprised a large subpopulation for which there was a dearth of information. Thus, eligible cases within each subgroup were randomly selected during each month of the study period in the following predefined percentages: all (100%) cases of patients who transferred to the ICU; one-third (33.3%) of cases of patients who expired; one-tenth (10%) of cases of patients with complex clinical events; and one-thirtieth (3.3%) of cases of patients without any of the aforementioned events. The number of cases sampled in any given month varied slightly due to different number of eligible cases in that month (ie, different monthly denominators for each subgroup). Over the full study period, our goal was that the percentages of cases sampled within each subgroup would approximate the target sampling percentage for that subgroup (see 'Sample size estimates').

Structured case record reviews

We employed NAM's definition of DEs as missed diagnostic opportunities during the hospital encounter. As previously reported, adjudicators (hospitalists, advanced practice providers) were trained to use the revised Safer Dx instrument and modified Diagnostic Error Evaluation and Research (DEER) Taxonomy adapted for acute care to judge the presence of DE and identify process failures, respectively.[2 12 24 25] Each adjudicator was required to review five training cases, complete a review form (online supplemental appendix A) and discuss each case with an expert hospitalist reviewer (AKD, JLS). Both expert reviewers were senior hospitalists with 20+ years of clinical experience and led development of the structured case review process used in this and concurrent studies.[12–14 20]

All sampled cases were sent for independent review by two trained adjudicators. Case information, including the primary diagnoses at admission and discharge, and secondary diagnoses were abstracted. Admission notes, discharge summaries, consultant notes, escalation events (rapid response, code), nursing documentation preceding an escalation event, and objective data such as vital signs, medications orders and laboratory results were reviewed. Reviewers assessed the likelihood of DE independently of harm assessment.[24] Using Likert ratings, harm severity was categorised as minor (mild symptoms, short-term loss of function, minimal intervention), moderate (symptomatic, requiring intervention, increased length of stay, long-term loss of function), major (symptomatic, requiring life-saving, surgical or medical intervention, shortening life expectancy, permanent loss of function) or fatal; and harm preventability was classified as definitely not, probably not, probably or definitely.[2 12 13 24 26 27]

Using the modified DEER Taxonomy which we previously adapted for acute care,[2] adjudicators identified any of 44 process failures in nine diagnostic domains: access and presentation, history, physical exam, assessment, diagnostic testing, diagnostic information and follow-up, subspecialty consultation, team communication and collaboration, and patient experience (online supplemental appendix A).[12] For example, to attribute a diagnostic process failure related to the patient experience, the adjudicator would use available clinical documentation to judge whether the patient had received an accurate and timely explanation of their health problem; whether there was a delay in communicating test results, assessments or consultant findings; or whether the care team did not address patient concerns, preferences or non-adherence.[22]

Reviewers met virtually to resolve discrepancies and complete the consensus review form while re-reviewing the case in the EHR. Major deterioration events (eg, rapid responses, increasing oxygen requirement) and relevant laboratory tests and orders (eg, blood cultures, antibiotics) were reviewed to identify delays in care (eg, ordering lactate levels within accepted timeframes for sepsis). When available, autopsy findings were reviewed. For cases in which the diagnosis was uncertain at discharge, a DE designation was borderline or multiple process failures were selected, reviewers "looked forward" at emergency department visits, readmissions, unexpected surgeries or deaths up to 90 days after admission to assess whether a missed opportunity during index hospitalisation was responsible.[28] All DE descriptions were characterised as missed, incorrect or delayed.

In the validation study, we determined that while independent reviews yielded moderate agreement (typical for AE studies),[29] the consensus and expert panel conducted tertiary reviews yielded strong agreement (Cohen's kappa>0.7).[12] Thus, to ensure consistency in consensus reviews, all cases that had discrepancies in DE designation between reviewers; inconsistencies in DE description (did not meet the operational definition); multiple Safer Dx items that were borderline (slightly agree or disagree); multiple process failures selected; or an uncertain diagnosis (no clear aetiology) were flagged for tertiary review during the consensus review between independent reviewers. An expert panel composed of five clinician adjudicators (AKD, JLS, ABG, SR, DM) met bimonthly to review individual and consensus forms, pertinent EHR data, and final DE designation and descriptions for all cases flagged for tertiary review. To participate on the expert panel, clinician adjudicators had to independently review and discuss at least 25 cases with an expert reviewer (AKD, JLS).

The primary outcome was defined as the presence of a DE judged to have caused harm of any severity up to 90 days after admission. Secondary outcomes included preventable DEs and severely harmful DEs (having major or fatal harm). Demographic variables included age, sex, race, ethnicity, insurance, risk cohort and International Classification of Diseases, 10th Revision (ICD-10) problem group for admission

diagnoses. Independent variables were defined as the presence of process failures overall and within each diagnostic domain. Other measures included the type of diagnosis (primary or secondary) and ICD-10 diagnosis code associated with each harmful DE, and characterisation of DE as missed, incorrect or delayed.

Statistical analysis

Demographic characteristics, harmful DE rates within each of the four subgroups, and characteristics of harmful DEs were reported as numbers and percentages for categorical variables and means (standard deviations) for continuous variables as appropriate. Cohen's kappa was calculated to assess inter-rater reliability between individual reviews, and between final consensus and expert panel tertiary reviews.[12]

For the primary analysis, we oversampled cases with high-risk criteria (figure 1) to gain more information on the probability of DE to obtain weighted estimates of outcomes in the population. To calculate unbiased estimates of harmful DEs, each case was weighted by the inverse probability of being sampled from its subgroup.[30] Thus, the study can be considered a stratified complex survey (one strata for each of the four subgroups) with weighting defined as the inverse probability of being sampled. Complex survey weighting for stratified samples was applied to obtain unbiased estimates of population characteristics including standard errors and 95% confidence intervals (95% CIs). Complex survey weighting was similarly applied to estimate DEs, harmful DE and preventable DE rates within subgroups (eg, age, sex, race, ethnic group, insurance, risk cohort).[31] Thus, the weighted stratified sampling design offered flexibility in oversampling certain subgroups, while allowing for complex survey methods to reweight the observations in the sample to obtain unbiased estimates of any population characteristic.

For secondary analyses, we used multivariable logistic regression to model the primary outcome using the presence of process failures within each diagnostic domain as independent variables to understand the extent to which each domain was associated with harmful DEs. Cases were weighted by their inverse probability of being sampled, and standard errors and 95% CIs from the logistic regression were calculated using complex survey weighting for stratified samples.[32 33] Because assessment failures frequently contribute to DE, we developed two models, one with and one without assessment failures.[14] All analyses were performed using complex survey procedures in SAS version 9.4 (SAS Institute).

Sample size estimate

To obtain an estimate and CI for the percentage of hospitalised patients with a harmful DE, we based our required sample size on having a 95% weighted binomial complex survey CI that was at most 4% wide. Thus, we required at least 700 patients overall to ensure that the resulting 95% CI would be ±2% (4% wide) of the estimated percentage. To ensure a sample size of at least 100 cases within each subgroup while oversampling cases from the complex clinical event subgroup relative to others, we aimed for a 95% binomial CI that would be ±8% (16% wide) of the estimated percentage for each subgroup.

RESULTS

Of the 9147 eligible cases (figure 1), 675 were randomly sampled from each subgroup during each month of the study period in the following percentages: 130 (100%) ICU transfers after 24 hours, 141 (38.5%) cases of patients who expired within 90 days of admission, 298 (7%) cases with complex clinical events and 106 (2.4%) cases without the aforementioned high-risk criteria. Demographics of eligible and sampled cases are reported in table 1. The values for Cohen's kappa for DE determination were 0.52 between individual reviewers and 0.87 between final consensus and expert panel tertiary reviews. By individual subgroup, harmful DEs were identified in 37 cases of ICU transfers after 24 hours (28.5%, 95% CI 20.70 to 36.22), 18 cases of deaths within 90 days of admission (12.8%, 95% CI 7.26 to 18.27), 23 cases with complex clinical events (7.7%, 95% CI 4.69 to 10.75) and 6 cases with no events (5.7%, 95% CI 1.26 to 10.06).

The population estimates of harmful, preventable and severely harmful DEs (table 2) based on the weighted sample were 7.2% (95% CI 4.66 to 9.80), 6.1% (95% CI 3.79 to 8.50) and 1.1% (95% CI 0.55 to 1.68), respectively. Based on these estimates, 84.7% of harmful DEs were preventable. Harmful DE estimates were higher for older, White, non-Hispanic, non-privately insured and high-risk patients.

The presence of process failures was significantly associated with harmful DEs when assessment failures were included (1.74, 95% CI 1.45 to 2.09, $p<0.01$) and excluded (1.94, 95% CI 1.62 to 2.33, $p<0.01$). In multivariable model 1 (table 3), assessment failures had the largest association with harmful DEs (7.34, 95% CI 3.86 to 13.95, $p<0.01$). In multivariable model 2 without assessment failures (table 3), harmful DEs were significantly associated with failures in diagnostic testing (odds ratio (OR) 4.24), subspecialty consultation (OR 3.11), patient experience (OR 2.93) and history (OR 2.50).

The severity of harm attributed to DEs experienced by 84 patients (table 4) was characterised as minor in 5 (6.0%), moderate in 36 (42.9%), major in 25 (29.8%) and fatal in 18 (21.4%). Forty (47.6%) were related to the primary diagnosis at admission or discharge and 44 (52.4%) were related to a secondary diagnosis. Fifty-two (61.9%) were characterised as delays. Errors associated with major or fatal harm were frequent in the high-risk cohort (55.1%, 43/78) and infrequent in

Table 1 Demographic characteristics of eligible cases, random sample and population based on weighted study sample

Characteristic	Eligible (n=9147)	Random sample (n=675)	Population based on weighted study sample*
Unique patients, n (%)	7390	649	–
Age in years, mean (SD)	61.6 (18.7)	65.8 (18.1)	62.4 (60.5 to 64.3)
Female, n (%)	5105 (55.8)	370 (54.8)	58.1 (52.8 to 63.2)
Race/ethnicity, n (%)			
White non-Hispanic	5264 (57.6)	416 (61.6)	60.4 (55.1 to 65.5)
Black non-Hispanic	1580 (17.3)	109 (16.2)	17.3 (13.6 to 21.8)
Hispanic	1522 (16.6)	103 (15.3)	16.5 (12.9 to 20.9)
Other or missing	781 (8.5)	47 (7.0)	5.8 (3.8 to 8.7)
Marital status, n (%)			
Married or partnered	3582 (39.2)	287 (42.5)	41.1 (36.0 to 46.4)
Separated, divorced or widowed	2064 (22.6)	173 (25.6)	23.0 (18.9 to 27.6)
Single never married	3446 (37.7)	210 (31.1)	35.6 (30.7 to 40.9)
Unknown	55 (0.6)	5 (0.7)	0.3 (0.1 to 1.1)
Primary language English, n (%)	8039 (87.9)	588 (87.1)	84.4 (79.9 to 88.0)
Education, n (%)			
Less than high school	1114 (12.2)	97 (14.4)	15.8 (12.2 to 20.3)
Graduated high school or GED	3087 (33.8)	228 (33.8)	31.6 (26.9 to 36.8)
Some college	983 (10.8)	69 (10.2)	12.4 (9.3 16.5)
Graduated college or higher education	3262 (35.7)	223 (33.0)	32.7 (27.9 to 37.7)
Unknown	701 (7.7)	58 (8.6)	7.4 (5.3 to 10.4)
Socioeconomic status (median income by zip code), n (%)			
≤$64 841	2270 (24.8)	129 (19.1)	21.2 (17.1 to 26.0)
$64 842–$98 333	2280 (24.9)	174 (25.8)	27.4 (22.8 to 32.4)
$98 334–$129 916	2330 (25.5)	201 (29.8)	27.8 (23.4 to 32.7)
>$129 916	2173 (23.8)	165 (24.4)	23.6 (19.4 to 28.4)
Insurance status, n (%)			
Commercial	3779 (41.3)	262 (38.8)	42.3 (37.1 to 47.6)
Medicaid	1246 (13.6)	80 (11.9)	14.2 (10.7 to 18.5)
Medicare	3999 (43.7)	323 (47.9)	42.7 (37.5 to 47.9)
Other	123 (1.3)	10 (1.5)	0.9 (0.34 to 2.4)
Primary care physician, n (%)			
Network	4679 (51.2)	355 (52.6)	53.9 (48.6 to 59.1)
Non-network	4468 (48.9)	320 (47.4)	46.1 (40.9 to 51.4)
Employment, n (%)			
Employed	2270 (24.8)	144 (21.3)	25.8 (21.3 to 30.8)
Unemployed	3248 (35.5)	215 (31.9)	34.8 (29.9 to 40.0)
Retired	3464 (37.9)	304 (45.0)	37.7 (32.8 to 43.0)
Unknown	165 (1.8)	12 (1.8)	1.7 (0.9 to 3.4)
Van Walraven score, mean (SD)	9.3 (9.4)	13.5 (10.8)	9.4 (8.5 to 10.4)
Readmission risk, mean (SD)	17.3 (12.3)	19.3 (13.1)	17.2 (16.1 to 18.3)
Deterioration index, mean (SD)	28.5 (11.7)	34.0 (14.3)	29.5 (28.0 to 31.0)
Cohort subtype			
High-risk†	4750 (51.9)	569 (84.3)	51.9 (48.2 to 55.7)
Low-risk‡	4397 (48.1)	106 (15.7)	48.1 (44.3 to 51.8)
ICD-10 problem group for admission diagnosis			
Blood and blood-forming organs (D50–D89)	220 (2.4)	13 (1.9)	2.0 (0.9 to 4.0)
Circulatory system (I00–I99)	617 (6.8)	57 (8.4)	8.3 (5.8 to 11.8)
Digestive system (K00–K95)	1495 (16.3)	110 (16.3)	18.7 (15.0 to 23.2)
Endocrine, nutritional and metabolic diseases (E00–E89)	663 (7.3)	52 (7.7)	7.1 (4.9 to 10.2)
Infectious and parasitic diseases (A00–B99)	125 (1.4)	6 (0.9)	0.4 (0.1 to 1.2)
Injuries (S00–T88)	219 (2.4)	16 (2.4)	1.9 (0.9 to 3.8)

Continued

Table 1 Continued

Characteristic	Eligible (n=9147)	Random sample (n=675)	Population based on weighted study sample*
Genitourinary system (N00–N99)	645 (7.1)	37 (5.5)	6.2 (4.0 to 9.6)
Mental, behavioural and neurodevelopmental disorders (F01–F99)	223 (2.4)	11 (1.6)	1.8 (0.8 to 4.2)
Musculoskeletal system and connective tissue (M00–M99)	383 (4.2)	22 (3.3)	3.7 (2.2 to 6.4)
Nervous system (G00–G99)	80 (0.9)	7 (1.0)	0.7 (0.2 to 2.6)
Neoplasms (C00–D49)	82 (0.9)	7 (1.0)	0.3 (0.1 to 1.1)
Respiratory system (J00–J99)	695 (7.6)	71 (10.5)	9.8 (7.1 to 13.3)
Skin and subcutaneous tissue (L00–L99)	287 (3.1)	12 (1.8)	2.9 (1.5 to 5.6)
Symptoms, signs and ill-defined conditions (R00–R99)	2821 (30.8)	196 (29.0)	27.6 (23.2 to 32.6)
Other§	276 (3.0)	25 (3.7)	5.0 (3.1 to 8.1)
Missing	316 (3.5)	33 (4.9)	3.5 (2.0 to 5.9)

*Population characteristic estimated based on weighted study sample, reported as percentage or mean (95% CI).
†High-risk cases include patients who transferred to the ICU transfer 24 hours or more after admission (including patients who expired after ICU transfer); expired within 90 days of admission (excluding patients who expired after ICU transfer); and had complex clinical events but did not transfer to the ICU or expire.
‡Low-risk cases include patients without any of the high-risk criteria.
§Eye and adnexa (H00–H59); ear and mastoid process (H60–H95); pregnancy (O00–O9A); congenital malformations, deformations, chromosomal abnormalities (Q00–Q99); special purposes (U00–U85); health status and services (Z00–Z99); morbidity (V00–Y99).
GED, General Educational Development; ICD-10, International Classification of Diseases, 10th Revision; ICU, intensive care unit; SD, standard deviation.

the low-risk cohort (0%, 0/6). The most frequent diagnoses (ICD-10) associated with these events included heart failure (I50.X), acute kidney failure (N17.9), sepsis (A41.X), pneumonia (J18.9), respiratory failure (J96.X), altered mental status (R41.82), abdominal pain (R10.9) and hypoxaemia (R09.02). Examples of harmful DEs, primary and secondary diagnoses, harm severity and preventability, and process failures are provided in online supplemental appendix B for selected cases in each of the four subgroups.

Table 2 Population estimates (n=9147) of harmful, preventable and severely harmful diagnostic errors for general medicine patients based on weighted sample (n=675)

Variable	Harmful DE	Harmful DE, preventable	Harmful DE, severe
Overall, n (%) (95% CI)	**662 (7.2) (4.66 to 9.80)**	**562 (6.1) (3.79 to 8.50)**	**102 (1.1) (0.55 to 1.68)**
Age in years			
<65	221 (4.5) (1.67 to 7.39)	161 (3.3) (1.10 to 5.50)	22 (0.5) (0.00 to 1.03)
>65	441 (10.3) (5.93 to 14.72)	401 (9.4) (5.09 to 13.71)	79 (1.9) (0.82 to 2.90)
Sex			
Female	363 (6.8) (3.40 to 10.29)	300 (5.6) (2.60 to 8.69)	58 (1.1) (0.30 to 1.88)
Male	298 (7.8) (3.92 to 11.62)	262 (6.8) (3.12 to 10.54)	44 (1.1) (0.34 to 1.96)
Race			
White	509 (9.0) (5.48 to 12.54)	483 (8.5) (5.05 to 12.04)	63 (1.1) (0.36 to 1.88)
Non-White	153 (4.4) (0.84 to 7.90)	79 (2.3) (0.00 to 4.61)	39 (1.1) (0.24 to 1.96)
Ethnic group			
Hispanic	6 (0.4) (0.01 to 0.84)	5 (0.4) (0.00 to 0.76)	5 (0.4) (0.00 to 0.76)
Non-Hispanic	655 (8.6) (5.52 to 11.64)	557 (7.3) (4.49 to 10.09)	39 (1.3) (0.59 to 1.94)
Insurance			
Private	148 (3.8) (1.68 to 5.97)	116 (3.0) (1.13 to 4.88)	32 (0.8) (0.06 to 1.59)
Public	509 (9.8) (5.60 to 13.90)	443 (8.5) (4.66 to 12.34)	67 (1.3) (0.48 to 2.11)
Other	5 (8.2) (0.00 to 22.85)	2 (3.8) (0.00 to 12.53)	2 (3.8) (0.00 to 12.53)
Risk cohort			
High-risk*	413 (8.7) (5.93 to 11.47)	358 (7.5) (4.94 to 10.12)	102 (2.1) (1.06 to 3.23)
Low-risk†	248 (5.6) (1.22 to 10.07)	204 (4.6) (0.63 to 8.66)	–

*High-risk cases include patients who transferred to the ICU transfer 24 hours or more after admission (including patients who expired after ICU transfer); expired within 90 days of admission (excluding patients who expired after ICU transfer); and had complex clinical events but did not transfer to the ICU or expire.
†Low-risk cases include patients without any of the high-risk criteria.
DE, diagnostic error; ICU, intensive care unit.

Table 3 Diagnostic process failures associated with harmful diagnostic errors: multivariable logistic regression models

Diagnostic domain	Model 1: Adjusted, weighted OR (95% CI)	P value	Model 2: Adjusted, weighted OR (95% CI)	P-value
Access and presentation	0.64 (0.25 to 1.67)	0.37	0.61 (0.20 to 1.84)	0.38
History	0.96 (0.44 to 2.06)	0.91	2.50 (1.29 to 4.84)	**<0.01**
Physical exam	0.63 (0.26 to 1.51)	0.30	0.79 (0.30 to 2.08)	0.64
Assessment	7.34 (3.86 to 13.95)	**<0.01**	–	–
Diagnostic test ordering, performance and interpretation	3.13 (1.68 to 5.82)	**<0.01**	4.24 (2.50 to 7.20)	**<0.01**
Diagnostic information and patient follow-up	0.95 (0.57 to 1.58)	0.83	1.29 (0.71 to 2.33)	0.40
Subspecialty consultation and referral	2.24 (1.10 to 4.54)	0.03	3.11 (1.40 to 6.89)	**<0.01**
Healthcare team communication and collaboration	1.15 (0.43 to 3.04)	0.79	1.74 (0.61 to 4.95)	0.30
Patient experience	3.51 (1.51 to 8.17)	<0.01	2.93 (1.43 to 5.98)	**<0.01**

Model 1: All diagnostic process domains including assessment failures.
Model 2: All diagnostic process domains excluding assessment failures.
Bold type denotes statistical significance.

DISCUSSION

We evaluated a weighted random sample of patients hospitalised on general medicine and estimate that about 1 in 14 patients (~7%) in this population experienced a harmful DE related to the primary diagnosis at either admission or discharge, and an equivalent percentage of secondary diagnoses. The majority of these harmful DEs were judged to be preventable. In multivariable analysis excluding assessment failures, failures in diagnostic testing, subspecialty consultation, patient experience, and history were associated with harmful DEs. These data suggest that DEs are frequent on general medicine, associated with certain process failures, and cause substantial harm.

While our observed severely harmful DE estimate (cases with major or fatal outcomes) of 1% is consistent with prior studies,[3 6] our overall estimate (including cases with less severe impact) was higher. Gunderson *et al* estimated that the incidence of harmful DEs in inpatients to be *at least* 0.7%.[3] This systematic review generated an estimate based on harmful DEs pooled from retrospective studies of enriched cohorts such as autopsy studies, many of which used an AE screening process similar to the Harvard Medical Malpractice Study and the recent study by Bates *et al*.[6 34] Experts have suggested that such methods are not well-suited to detect DEs.[6 7] Conversely, studies that screened for DEs by rigorously evaluating the diagnostic process have yielded higher event rates (5–18%) in cohorts of patients who expired or transferred to ICU or expired in the hospital, were under investigation for COVID-19 or were readmitted.[8 10 20]

Unlike prior studies that screened for DEs, our estimate reflects harmful DE rates related to exposure to hospital care received on the general medicine service, not limited to specific or enriched cohorts.[4 13 20 35 36] For example, the recent multicentre study by Auerbach *et al* (which was based on our approach) observed that harmful DEs occurred in 26% of patients who transferred to the ICU 24 hours or more after admission.[13 14] While our weighted sample included patients who transferred to the ICU and observed a similar rate (28.5%), we also sampled cases without these high-risk events to obtain a population estimate for hospitalised patients who received care on the general medicine service. By querying the EHR using clinical screening criteria and ensuring adequate sampling of each subgroup,[12 23] our sample broadly represented clinical trajectories typically encountered for hospitalised patients who received general medical care and was not limited to a specific disease process (eg, epidural abscess, myocardial infarction).[20 35] As might be expected, in cases with a major deterioration event, the harm was frequently characterised as major or fatal. In contrast, in cases without such events, the harm was frequently characterised as mild or moderate. Yet, the impact of the harmful DE did not necessarily correlate with the event. For example, for cases in which the patient expired after hospitalisation, the harmful DE was not always associated with the outcome as illustrated in Case 2 (online supplemental appendix B).

Additionally, many of the harmful DEs identified were frequently related to a secondary diagnosis or a diagnostic delay (ie, a missed diagnostic opportunity early during hospital encounter). In Case 2, the harmful DE was associated with an undetected pleural effusion, a secondary diagnosis that was missed during the index hospitalisation but identified during subsequent readmission for hepatic hydrothorax. In Case 3, the harmful DE was related to both the primary diagnosis (sepsis) and secondary diagnosis (methicillin-susceptible *Staphylococcus aureus* bacteremia); however, identifying pelvic abscess as the source of bacteremia was identified late in the hospital course after obtaining dedicated imaging.

Our ability to detect harmful DEs beyond those captured by traditional methods such as the IHI Global Trigger tool can be explained by our structured case review process that empowered adjudicators to use the EHR to *rigorously* assess the diagnostic process and consider the impact of identified DEs both during the course of hospitalisation and afterwards for cases

Table 4 Characteristics of harmful diagnostic errors, harm severity and International Classification of Diseases,10th Revision (ICD-10) codes by risk cohort (n=84)

Cohort	ICU transfer after 24 hours (n=37)	Death within 90 days (n=18)	Complex clinical events (n=23)	None (n=6)
Diagnosis associated with harmful DE, n (%)				
Primary diagnosis at admission or discharge	19 (51.4)	8 (44.4)	9 (39.1)	4 (66.7)
Secondary diagnosis	18 (48.6)	10 (55.6)	14 (60.9)	2 (33.3)
Characterisation of harmful DE, n (%)				
Missed diagnosis	6 (16.2)	8 (44.4)	9 (39.1)	4 (66.7)
Incorrect diagnosis	4 (10.8)	–	1 (4.4)	–
Delayed diagnosis	27 (73.0)	10 (55.6)	13 (56.5)	2 (33.3)
Harm severity, n (%)				
Minor (symptomatic, symptoms are mild, loss of function or harm is minimal or intermediate but short-term, and no or minimal intervention is required)	–	1 (5.6)	3 (13.0)	1 (16.7)
Moderate (symptomatic, requiring intervention, an increased LOS, or causing permanent or long-term harm or loss of function)	9 (37.5)	5 (27.8)	17 (73.9)	5 (83.3)
Major (symptomatic, requiring life-saving intervention or major surgical or medical intervention, shortening life expectancy or causing major permanent or long-term harm or loss of function)	16 (43.2)	7 (38.9)	2 (8.7)	–
Fatal (on balance of probabilities, death was caused or brought forward in the short term by the incident)	12 (32.4)	5 (27.8)	1 (4.3)	–
*ICD-10 problem group (most frequent diagnoses), n (%)				
Blood and blood-forming organs (anaemia, D64.9, n=2)	–	2 (11.1)	1 (4.4)	–
Circulatory system (heart failure, I50.X, n=8)	8 (21.6)	1 (5.6)	3 (13.0)	1 (16.7)
Digestive system (acute pancreatitis, K85.XX, n=2)	2 (5.4)	1 (5.6)	3 (13.0)	0 (0)
Endocrine, nutritional and metabolic diseases (type 1 diabetes mellitus with ketoacidosis, E10.1, n=2)	3 (8.1)	2 (11.1)	1 (4.4)	1 (16.7)
Infectious and parasitic diseases (sepsis, A41.X, n=3)	1 (2.7)	2 (11.1)	3 (13.0)	–
Injuries (intracranial injury, S06.X; minor contusion of kidney, S37.019A; trochanteric fracture of right femur, S72.101)	1 (2.7)	1 (5.6)	1 (4.4)	–
Genitourinary system (acute kidney failure, N17.9, n=4)	1 (2.7)	3 (16.7)	2 (8.7)	–
Mental, behavioural and neurodevelopmental disorders (alcohol dependence with withdrawal, F10.23; sedative, hypnotic or anxiolytic dependence and withdrawal, F13.239)	2 (5.4)	–	–	–
Musculoskeletal system and connective tissue (osteomyelitis, M86.9; microscopic polyangiitis, M31.7; spinal stenosis, M48.02; pain in unspecified shoulder, M25.519)	–	1 (5.6)	3 (13.0)	–
Neoplasm	–	–	–	–
Nervous system (extradural and subdural abscess, G06.2)	1 (2.7)	–	–	–
Respiratory system (pneumonia, J18.9, n=4; respiratory failure, J96.X, n=3)	9 (24.3)	3 (16.7)	2 (8.7)	–
Skin and subcutaneous tissue	–	–	–	–
Symptoms, signs and ill-defined conditions (altered mental status, R41.82, n=4; abdominal pain, R10.9, n=3; hypoxaemia, R09.02, n=3)	9 (24.3)	2 (11.1)	4 (17.4)	4 (66.7)
Other*	–	–	–	–

*Eye and adnexa (H00–H59); ear and mastoid process (H60–H95); pregnancy (O00–O9A); congenital malformations, deformations, chromosomal abnormalities (Q00–Q99); special purposes (U00–U85); health status and services (Z00–Z99); morbidity (V00–Y99).
DE, diagnostic error; ICD-10, International Classification of Diseases, 10th Revision; ICU, intensive care unit; LOS, length of stay.

with uncertainty or multiple process failures.[12 27] This approach enabled detection of events with both severe and less severe outcomes, such as a delayed diagnosis of pelvic abscess in Case 3 (online supplemental appendix B) in which the patient did not expire or transfer to the ICU. Unlike studies focused on sampling highest-risk events,[14] our inclusion of subgroups with complex clinical events and no events (large and understudied subpopulations) and use of targeted reviews of post-hospitalisation documentation using a "look forward" approach[28] generated new insights about the spectrum of harms associated with faulty diagnostic processes during the hospital encounter. For example, DEs were judged to be present in cases in which the diagnosis was uncertain (unclear aetiology of altered mental status) and specific process failures (misinterpretation of electroencephalogram results documented at discharge in relation to final report, a caregiver-reported concern about being discharged too soon) were identified from review of an unanticipated event after index

hospitalisation (readmission for seizure for initiation of anticonvulsants) in Case 4 (online supplemental appendix B). Interestingly, despite the heterogeneity of patient experiences, certain patient- or caregiver-reported concerns when documented could serve as important clues about a faulty diagnostic process.[2 15]

Our multivariable analyses suggest that certain process failures are frequently associated with harmful DEs. These include uncertainty in initial assessments, complex diagnostic testing and interpretation, suboptimal subspecialty consultation, patient-reported concerns and history-taking. A thorough analysis of such events should yield insights for optimising triggers for surveillance[3 37–39] and developing preventative interventions.[2 17] For example, triggers or interventions could be developed by analysing the timing, sequence and pattern of consultation or test orders in EHR audit logs,[40 41] capturing diagnostic concerns from patients who review online notes using a patient diagnostic questionnaire[15 42] and applying machine learning and natural language processing to model uncertainty expressed in documentation.[43–46]

Additionally, the rapid adoption of artificial intelligence (AI) and large language models (GPT-4) has much potential to facilitate prospective case surveillance by detecting complex patterns of risk factors and clinical events that represent markers of risk or suboptimal diagnostic processes. For example, once trained on large cohorts and inclusive of data retrieved from various sources (EHR, institutional safety reporting systems, patients), AI-based tools could facilitate detection of diagnostic uncertainty in initial assessments; complex sequences of diagnostic tests; incorrect study interpretations; discrepancies in consultation recommendations; patient–clinician diagnostic discordance; patient-reported diagnostic concerns; or lack of improvement based on expected clinical course. Furthermore, when embedded in the EHR and integrated into workflow for clinicians, real-time AI-generated insights and diagnostic suggestions could prompt more timely intervention, such as pausing to take a diagnostic time-out and reconsidering the working diagnosis as we recently described.[17 47–49]

While our study has several strengths, it has limitations. First, it was conducted using a non-traditional stratified sampling approach at a single institution for patients who received general medical care and had a length of stay <21 days. While our sampling approach was grounded in emerging research, expert consensus and local data, event rates may differ for patients who receive more specialised care delivered on other services and at other institutions, and are likely higher for patients with longer exposure to hospital care. Regarding top disease categories implicated in serious harms from DEs, while we frequently observed harmful DEs related to infection (sepsis, pneumonia), we infrequently identified harmful DEs related to vascular or cancer diagnoses.[4] Such cases may be detected in patients who receive specialised care delivered on cardiovascular and oncology services, not a general medicine service. Second, for reasons indicated earlier, we excluded cases during the initial waves of the pandemic including patients hospitalised on COVID-19 teams. While unintended bias is possible, recent data suggest that harmful DE rates are similar in this population.[20]

Third, we used the EHR (known to contain inaccurate information about the status of death) to identify patients who expired within 90 days of admission.[50] Furthermore, while we did not consider post-discharge events such as readmissions in our queries to stratify lower-risk subgroups, in about one-third of our study sample we relied on a "look forward" approach to identify these events when individual reviewers did not agree on DE determination, when there was uncertainty about the diagnosis, or when multiple process failures where present. Future studies should consider such factors when defining criteria for subgroups.

Fourth, limiting the period of harm to 90 days from admission may have precluded detection of certain serious DEs with more delayed impact (such as lack of follow-up of incidental pulmonary nodules); however, other safety net systems likely mitigated such faulty diagnostic processes at our institution.[51] Lastly, we observed moderate inter-rater reliability between individual reviews. As in our validation study,[12] we observed substantial agreement between final consensus reviews and expert panel tertiary reviews, suggesting that the most important step is "talking through" determination of DEs and characterisation of associated harms for independently reviewed cases.

In summary, we performed a single-centre evaluation to estimate the prevalence of harmful DEs in hospitalised patients who received general medical care. While our results and sampling approach should be validated in larger samples, for different clinical services and at other sites, these data offer direction for improving surveillance approaches and developing preventative interventions. Novel approaches, including the use of AI and machine learning, have potential for facilitating more granular surveillance in large subpopulations without highest-risk events (such as the complex clinical events subgroup) than can be achieved by human review of the EHR alone; assessing uncertainty or risk in diagnostic processes[43]; and prompting preventative intervention to promote a culture of diagnostic safety.[6 7 47]

Author affiliations
[1]Department of Medicine, Division of General Internal Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[2]Harvard Medical School, Boston, Massachusetts, USA
[3]Mass General Brigham, Boston, Massachusetts, USA
[4]Vanderbilt University Medical Center, Nashville, Tennessee, USA
[5]Dartmouth-Hitchcock Medical Center, Lebanon, Pennsylvania, USA
[6]Columbia University Vagelos College of Physicians and Surgeons, New York, New York, USA

[7]Department of Industrial Engineering, Northeastern University - Boston Campus, Boston, Massachusetts, USA

X Anuj K Dalal @tweet_akdMD, Jeffrey L Schnipper @drjschnip and David W Bates @dbatessafety

Collaborators Patient Safety Learning Laboratory Adjudicator Group: David Lee MD; Daniel Palazuelos MD, MPH; Myrna Katalina Serna MD, MPH; Anne Kozak MS, PA-C; Khelsea O'Brien MS, PA-C; Shela Shah MD; Mohammed Wazir MD; Chadi Cortas MD, MD, MBA; Caroline Yang MD.

Contributors All authors contributed sufficiently to the conceptualisation (AKD, PG, RR, JG, JLS, SL, DWB), methodology (AKD, SP, DM-C, PG, RR, JG, JLS, SL, DWB); data curation (SP, KK, DM-C, AG, SR, MM, AL, NP, JMF); analysis (AKD, SP, KK, DM-C, MM, NP, CY, JMF, JLS, SL, DWB); project administration (AKD, SP, KK, DM-C, MM, AG, SR, AL, NP, ML); writing, editing and review of the manuscript (AKD, SP, KK, DM-C, AG, SR, MM, AG, AL, NP, ML, CY, JMF, PG, KOS, RR, JG, JLS, SL, DWB); and/or supervision and acquisition of funding (AKD, DWB).The guarantor (AKD) accepts full responsibility for the conduct of the study, analysis and access to data, decision to publish, and the finished work product.

Funding This study was funded by Agency for Healthcare Research and Quality (AHRQ) (R18 HS026613).

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Anuj K Dalal http://orcid.org/0000-0003-1431-2972
Abhishek Goyal http://orcid.org/0000-0001-7745-886X
Kumiko O Schnock http://orcid.org/0000-0002-1107-6399
Jeffrey L Schnipper http://orcid.org/0000-0001-7072-0781
David W Bates http://orcid.org/0000-0001-6268-1540

REFERENCES

1 Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*. Washington (DC): National Academies Press (US), 2015. Available: https://www.ncbi.nlm.nih.gov/books/NBK338596/
2 Griffin JA, Carr K, Bersani K, *et al*. Analyzing diagnostic errors in the acute setting: a process-driven approach. *Diagnosis (Berl)* 2022;9:77–88.
3 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;29:1008–18.
4 Newman-Toker DE, Nassery N, Schaffer AC, *et al*. Burden of serious harms from diagnostic error in the USA. *BMJ Qual Saf* 2024;33:109–20.
5 Classen DC, Lloyd RC, Provost L, *et al*. Development and evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf* 2008;4:169–77.
6 Bates DW, Levine DM, Salmasian H, *et al*. The safety of inpatient health care. *N Engl J Med* 2023;388:142–53.
7 Shojania KG. The elephant of patient safety: what you see depends on how you look. *Jt Comm J Qual Patient Saf* 2010;36:399–401.
8 Bergl PA, Taneja A, El-Kareh R, *et al*. Frequency, risk factors, causes, and consequences of diagnostic errors in critically ill medical patients: a retrospective cohort study. *Crit Care Med* 2019;47:e902–10.
9 Bergl PA, Zhou Y. Diagnostic error in the critically ill: a hidden epidemic? *Crit Care Clin* 2022;38:11–25.
10 Raffel KE, Kantor MA, Barish P, *et al*. Prevalence and characterisation of diagnostic error among 7-day all-cause hospital medicine readmissions: a retrospective cohort study. *BMJ Qual Saf* 2020;29:971–9.
11 Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. *BMJ Qual Saf* 2015;24:103–10.
12 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;9:446–57.
13 Dalal AK, Schnipper JL, Raffel K, *et al*. 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* 2024;19:140–5.
14 Auerbach AD, Lee TM, Hubbard CC, *et al*. Diagnostic errors in hospitalized adults who died or were transferred to intensive care. *JAMA Intern Med* 2024;184:164–73.
15 Giardina TD, Choi DT, Upadhyay DK, *et al*. Inviting patients to identify diagnostic concerns through structured evaluation of their online visit notes. *J Am Med Inform Assoc* 2022;29:1091–100.
16 de Vries EN, Ramrattan MA, Smorenburg SM, *et al*. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008;17:216–23.
17 Garber A, Garabedian P, Wu L, *et al*. Developing, pilot testing, and refining requirements for 3 EHR-integrated interventions to improve diagnostic safety in acute care: a user-centered approach. *JAMIA Open* 2023;6:ooad031.
18 Kazakova SV, Baggs J, Parra G, *et al*. Declines in the utilization of hospital-based care during COVID-19 pandemic. *J Hosp Med* 2022;17:984–9.
19 Cram P, Anderson ML, Shaughnessy EE. All hands on deck: learning to "un-specialize" in the COVID-19 pandemic. *J Hosp Med* 2020;15:314–5.
20 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;38:1902–10.
21 Gandhi TK, Singh H. Reducing the risk of diagnostic error in the COVID-19 era. *J Hosp Med* 2020;15:363–6.
22 Shojania KG, Burton EC, McDonald KM, *et al*. Changes in rates of autopsy-detected diagnostic errors over time. *JAMA* 2003;289:2849.
23 Shenvi EC, El-Kareh R. Clinical criteria to screen for inpatient diagnostic errors: a scoping review. *Diagnosis (Berl)* 2015;2:3–19.

24 Singh H, Khanna A, Spitzmueller C, *et al*. Recommendations for using the Revised Safer Dx Instrument to help measure and improve diagnostic safety. *Diagnosis (Berl)* 2019;6:315–23.
25 Schiff GD, Hasan O, Kim S, *et al*. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;169:1881–7.
26 Dubois RW, Brook RH. Preventable deaths: who, how often, and why? *Ann Intern Med* 1988;109:582–9.
27 Al-Mutairi A, Meyer AND, Thomas EJ, *et al*. Accuracy of the Safer Dx Instrument to identify diagnostic errors in primary care. *J Gen Intern Med* 2016;31:602–8.
28 Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. *BMJ Qual Saf* 2018;27:557–66.
29 Patient Safety Network. Patient safety 101. Secondary patient safety network. 2019. Available: https://psnet.ahrq.gov/primer/patient-safety-101
30 Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc* 1952;47:663.
31 Rao JNK, Scott AJ. The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J Am Stat Assoc* 1981;76:221–30.
32 Wilson JR, Lorenz KA. *Weighted Logistic Regression Model. Modeling Binary Correlated Responses Using SAS, SPSS and R*. Cham: Springer International Publishing, 2015:81–102.
33 Morel JG. Logistic regression under complex survey designs. *Surv Methodol* 1989;15:203–23.
34 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;324:377–84.
35 Newman-Toker DE, Wang Z, Zhu Y, *et al*. Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: toward a national incidence estimate using the “Big Three”. *Diagnosis (Berl)* 2021;8:67–84.
36 Winters B, Custer J, Galvagno SM Jr, *et al*. Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. *BMJ Qual Saf* 2012;21:894–902.
37 Skinner TR, Scott IA, Martin JH. Diagnostic errors in older patients: a systematic review of incidence and potential causes in seven prevalent diseases. *Int J Gen Med* 2016;9:137–46.
38 Murphy DR, Meyer AN, Sittig DF, *et al*. Application of electronic trigger tools to identify targets for improving diagnostic safety. *BMJ Qual Saf* 2019;28:151–9.
39 Bhise V, Sittig DF, Vaghani V, *et al*. An electronic trigger based on care escalation to identify preventable adverse events in hospitalised patients. *BMJ Qual Saf* 2018;27:241–6.
40 Adler-Milstein J, Adelman JS, Tai-Seale M, *et al*. EHR audit logs: a new goldmine for health services research? *J Biomed Inform* 2020;101:103343.
41 Kannampallil T, Adler-Milstein J. Using electronic health record audit log data for research: insights from early efforts. *J Am Med Inform Assoc* 2022;30:167–71.
42 Giardina TD, Korukonda S, Shahid U, *et al*. Use of patient complaints to identify diagnosis-related safety concerns: a mixed-method evaluation. *BMJ Qual Saf* 2021;30:996–1001.
43 Marshall TL, Nickels LC, Brady PW, *et al*. Developing a machine learning model to detect diagnostic uncertainty in clinical documentation. *J Hosp Med* 2023;18:405–12.
44 Bhise V, Rajan SS, Sittig DF, *et al*. Defining and measuring diagnostic uncertainty in medicine: a systematic review. *J Gen Intern Med* 2018;33:103–15.
45 Dahm MR, Crock C. Understanding and communicating uncertainty in achieving diagnostic excellence. *JAMA* 2022;327:1127–8.
46 Schnock KO, Garber A, Fraser H, *et al*. Providers’ and patients’ perspectives on diagnostic errors in the acute care setting. *Jt Comm J Qual Patient Saf* 2023;49:89–97.
47 Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. *JAMA* 2023;330:317–8.
48 Hirosawa T, Harada Y, Mizuta K, *et al*. Evaluating ChatGPT-4’s accuracy in identifying final diagnoses within differential diagnoses compared with those of physicians: experimental study for diagnostic cases. *JMIR Form Res* 2024;8:e59267.
49 Mizuta K, Hirosawa T, Harada Y, *et al*. Can ChatGPT-4 evaluate whether a differential diagnosis list contains the correct diagnosis as accurately as a physician? *Diagnosis (Berl)* 2024;11:321–4.
50 Wenger NS, Sanz Vidorreta FJ, Dudley MT, *et al*. Consequences of a health system not knowing which patients are deceased. *JAMA Intern Med* 2024;184:213–4.
51 Lacson R, O’Connor SD, Sahni VA, *et al*. Impact of an electronic alert notification system embedded in radiologists’ workflow on closed-loop communication of critical results: a time series analysis. *BMJ Qual Saf* 2016;25:518–24.

APPENDIX A: Final Chart Review Tool

Record ID _____

Patient MRN: [mrn]

Patient Name: [name]

Admission Date: [admit_date]

Chart Reviewer Name

Chart Review Date/Time _____

As you proceed through this patient's chart, please keep the following definition in mind:

A diagnostic error is defined as a missed opportunity to make an accurate or timely diagnosis (a missed, incorrect, or delayed diagnosis) related to the acute care episode

You may get a clue about a diagnostic error based on the clinical impression of various care team members during the course of the hospital encounter.

Chart Review Process:

1. Go into the patient's chart, open the "Encounters" tab, and find the hospital encounter of interest.
2. Read through the entire Discharge Summary to get an overview of what happened.
3. Open the patient's inpatient chart for the selected encounter and read through the entire Admission H&P (CC, HPI, ED course, Assessment and Plan, etc.) to understand the initial thought process and treatment plan.
4. Review objective data (use Event Log to review vitals, orders, EMAR, lab results, timing of treatments, consults, procedures, etc.)
5. Review subjective notes (i.e., the floor course documented by primary team, consults, nursing notes, ancillary staff, etc.) to identify discrepancies and clues early during the hospital course.
6. Consider whether a diagnostic error may have occurred during the hospital encounter based on actual clinical documentation.
7. Proceed to the chart review tool (next page).

Case Information

Did this patient go through the Emergency Department?
- ◯ Yes
- ◯ No

Where did this patient come from?
- ◯ Home
- ◯ Ambulatory Clinic (direct admit or referred by PCP or specialist)
- ◯ Skilled Nursing Facility / Rehab
- ◯ Transfer from an Outside Hospital or ED
- ◯ Other

Describe other _____

Was this patient admitted between the hours of 7 pm and 7 am (i.e., by the overnight team)?
- ◯ Yes
- ◯ No

*Can use 'Care Timeline' for time stamps

Brief Case Summary

*Can take from Brief Summary and Discharge Summary, and edit as you see fit based on your clinical judgment

_____

Diagnostic Timeline

Chief Complaint upon admission?

*Identify from CC or HPI in Admission H&P

_____

Upon ADMISSION, what did the primary team consider to be the primary diagnosis?

*Typically first problem listed in A&P by admitting team

_____

AT DISCHARGE, what did the primary team consider to be the primary diagnosis?

*Typically first problem of 'Hospital Course' in Discharge Summary

_____

Were the primary diagnoses upon admission and discharge the same?

*Different severity, stability, or acuity could make these diagnoses different (e.g., stable vs. unstable angina, acute on chronic kidney injury vs. chronic kidney injury)

◯ Yes
◯ No

Is the discharge diagnosis on the differential diagnosis list at admission?

◯ Yes
◯ No

Do you think there was a secondary diagnosis that was not appropriately addressed during the encounter?

For example, this could include a diagnosis that led to decompensation (RRT, code), transfer to or from another service, or that was present based on objective data but was not acknowledged in clinical documentation (drop in Hb/Hct from prior ambulatory value)

◯ Yes
◯ No

Please list the secondary diagnosis that was not appropriately addressed

_____

Please select a diagnosis which you will review.

◯ Primary admission diagnosis
◯ Primary discharge diagnosis
◯ Secondary diagnosis

The Safer Dx Instrument: Items for Determining The Likelihood of Diagnostic Error during the Hospital Encounter

Rate the following items for the ENITRE EPISODE OF CARE under review (i.e., from admission through discharge), related to the primary admission or discharge diagnosis, or a secondary diagnosis.

	Strongly Agree	Agree	Slightly Agree	Slightly Disagree	Disagree	Strongly Disagree
1. The documented history was suggestive of an alternate diagnosis, which was not considered to be the presumed or working diagnosis or was considered late.	○	○	○	○	○	○
2. The documented physical exam was suggestive of an alternate diagnosis, which was not considered to be the presumed or working diagnosis or was considered late.	○	○	○	○	○	○
3. Diagnostic testing data (laboratory, radiology, pathology or other results) were suggestive of an alternate diagnosis, which was not considered to be the presumed or working diagnosis or was considered late.	○	○	○	○	○	○
4. Data gathering through history, physical exam, and review of prior documentation (including prior laboratory, radiology, pathology or other results) was incomplete, given the patient's medical history and clinical presentation.	○	○	○	○	○	○
5. The diagnostic process was affected by incomplete or incorrect clinical information given to the care team by the patient or their primary caregiver.	○	○	○	○	○	○

6. The clinical information (i.e., history, physical exam or diagnostic data) should have prompted additional or earlier diagnostic evaluation through tests or consults.	○	○	○	○	○	○
7. The diagnostic reasoning was not appropriate, given the patient's medical history and clinical presentation.	○	○	○	○	○	○
8. Alarm symptoms or "Red Flags" (i.e., features in the clinical presentation that are considered to predict serious disease) were not acted upon in a timely manner.	○	○	○	○	○	○
9. Diagnostic data (laboratory, radiology, pathology or other results) available or documented were misinterpreted in relation to the subsequent final diagnosis.	○	○	○	○	○	○
10. There was missed or delayed follow-up of available diagnostic data (laboratory, radiology, pathology or other results) in relation to the subsequent final diagnosis.	○	○	○	○	○	○
11. The differential diagnosis was either not documented, OR the differential diagnosis documented did not include the subsequent final diagnosis.	○	○	○	○	○	○
12. The final diagnosis was not an evolution of the care team's initial presumed diagnosis (or working diagnosis).	○	○	○	○	○	○
13. The clinical presentation at the initial presentation was mostly typical of the final diagnosis for the hospital encounter.	○	○	○	○	○	○

In conclusion, based on all the above questions, the episode of care under review had a diagnostic error.

Diagnostic Error: A missed opportunity to make an accurate or timely diagnosis based on the available information, independent of harm (i.e., a missed, incorrect, or delayed diagnosis)

- ○ Strongly Agree
- ○ Agree
- ○ Slightly Agree
- ○ Slightly Disagree
- ○ Disagree
- ○ Strongly Disgree

Please include any notes regarding your final diagnostic error decision or individual Safer Dx tool decisions. (Optional)	_____
Please describe the diagnostic error Please be specific. For example, "delay in considering alternative etiologies of hypoxia after 4 days of appropriate IV abx for presumed bacterial pneumonia diagnosed on admission."	_____
Where did this error take place (check all that apply):	☐ Prior to admission to general medicine (e.g., ED, prior to transfer to medicine) ☐ While on general medicine ☐ After being on general medicine (e.g., after transfer to another unit)

Adverse Outcomes Related to the Episode of Care

Did the diagnostic error cause actual harm?	◯ Definitely ◯ Probably ◯ Probably Not ◯ Definitely Not
What is your confidence that the diagnostic error had potential to cause patient harm?	◯ Little or no confidence ◯ Slight confidence ◯ Less than 50-50 but close call ◯ More than 50-50 but close call ◯ Strong confidence ◯ Virtually certain confidence
What is the most likely severity of the diagnostic error's potential harm?	◯ Minor (Patient outcome is symptomatic, symptoms are mild, loss of function or harm is minimal or intermediate but short term, and no or minimal intervention is required.) ◯ Moderate (Patient outcome is symptomatic, requiring intervention, an increased LOS, or causing permanent or long term harm or loss of function.) ◯ Major (Patient outcome is symptomatic, requiring life-saving intervention or major surgical/medical intervention, shortening life expectancy or causing major permanent or long term harm or loss of function.) ◯ Death (On balance of probabilities; death was caused or brought forward in the short term by the incident.)

What was the severity of the diagnostic error's clinical impact?	◯ Minor (Patient outcome is symptomatic, symptoms are mild, loss of function or harm is minimal or intermediate but short term, and no or minimal intervention is required.) ◯ Moderate (Patient outcome is symptomatic, requiring intervention, an increased LOS, or causing permanent or long term harm or loss of function.) ◯ Major (Patient outcome is symptomatic, requiring life-saving intervention or major surgical/medical intervention, shortening life expectancy or causing major permanent or long term harm or loss of function.) ◯ Death (On balance of probabilities; death was caused or brought forward in the short term by the incident.)
Describe the actual clinical impact of the diagnostic error	_____
Describe the potential future clinical impact of the diagnostic error	_____
What is the likelihood that this clinical impact was preventable?	◯ Definitely not preventable ◯ Probably not preventable ◯ Probably preventable ◯ Definitely preventable
What is the likelihood that this clinical impact was ameliorable (i.e., whether the duration or severity of the harm could have been reduced or mitigated.)	◯ Definitely not ameliorable ◯ Probably not ameliorable ◯ Probably ameliorable ◯ Definitely ameliorable

Modified DEER Taxonomy Tool adapted for acute care

Please check any of the following DIAGNOSTIC PROCESS FAILURES that were present/occurred during the episode of care under review. Also, please identify those that had significant impact in causing the diagnostic error by checking "Significant".

Access/Presentation

	Present/Occurred	Significant
A. Failure or delay in patient seeking care	☐	☐
B. Failure or denial of access to care	☐	☐
C. Failure of triage or admission to wrong service	☐	☐

History

	Present/Occurred	Significant
A. Failure or delay in providing or eliciting a piece of history data	☐	☐
B. Inaccurate or misinterpreted piece of history data	☐	☐
C. Suboptimal weighing of a piece of history data	☐	☐
D. Failure or delay in acting on or following-up on a piece of history data	☐	☐

Physical Exam

	Present/Occurred	Significant
A. Failure or delay in eliciting critical physical examination finding	☐	☐
B. Inaccurate or misinterpreted physical examination finding	☐	☐
C. Suboptimal weighing of a physical examination finding	☐	☐
D. Failure or delay in acting on or following-up on a physical examination finding	☐	☐

Assessment

	Present/Occurred	Significant
Failure or delay in considering correct diagnosis	☐	☐
Suboptimal weighing or prioritizing of primary and/or secondary diagnose	☐	☐
Too much weight to lower probability/priority diagnosis	☐	☐

Diagnostic Test Ordering, Performance, and Interpretation

	Present/Occurred	Significant
A. Failure or delay in ordering needed test(s)	☐	☐
B. Failure or delay in performing needed test(s)	☐	☐
C. Suboptimal test sequencing	☐	☐

D. Failure to order correct test(s) (e.g., ordered head CT for suspected cerebellar stroke)	☐	☐
E. Failure to order test(s) in correct way	☐	☐
F. Identification failure (e.g., sample mix-up, mislabeled specimen, or test performed on the wrong patient)	☐	☐
G. Technical or processing error (equipment problem, poor processing of specimen/test, or skill issue)	☐	☐
H. Specimen delivery problem (e.g., specimen never sent, delayed delivery, or lost specimen)	☐	☐
I. Erroneous reading of test (lab/radiology)	☐	☐
J. Erroneous clinician interpretation of test	☐	☐

Diagnostic Information and Patient Follow-up

	Present/Occurred	Significant
A. Failure or delay in acting on or following-up on test result (including results not communicated to the patient)	☐	☐
B. Failure or delay to re-test (e.g., follow-up lactate, INR)	☐	☐
C. Failure or delay in monitoring (e.g., failure to routinely check vital signs, failure to apply monitor, technical issue)	☐	☐
D. Missed physiologic monitoring finding (e.g., persistent hypoxia, oxygen requirement)	☐	☐
E. Failure or delay in recognizing or acting upon urgent condition or complications	☐	☐
F. Failure to refer the patient to appropriate setting or for appropriate monitoring	☐	☐

G. Failure or delay in timely follow-up or re-examination of the patient	☐	☐

Subspecialty Consultation/Referral

	Present/Occurred	Significant
A. Failure or delay in ordering a referral or consult	☐	☐
B. Failure or delay in obtaining or scheduling an ordered referral or consult	☐	☐
C. Failure or delay of the consulting team to see the patient	☐	☐
D. Suboptimal consultation (e.g., error in diagnostic consultation performance) or follow-up of consultation	☐	☐
E. Inappropriate or unneeded referral or consultation	☐	☐

Healthcare Team Communication & Collaboration

	Present/Occurred	Significant
A. Failure or delay in communication of clinical assessment at initial and subsequent encounters between healthcare team members	☐	☐
B. Failure or delay in transmission or communication of lab/test result(s) to healthcare providers	☐	☐
C. Failure or delay in communication of critical information between pathologists, radiologists, or technologists and the primary team	☐	☐
D. Failure or delay in communication between consultants and the patient's primary team	☐	☐

E. Failure or delay in communication of critical information within the patient's primary team (e.g., a missed hand-off between the night and day teams or a lack of communication during rounds.) May include the patient's nurse, pharmacist, therapist, social worker, physician, etc.	☐	☐

Patient Experience

	Present/Occurred	Significant
A. Failure to communicate an accurate and timely explanation of the patient's health problem(s) to the patient/caregiver	☐	☐
B. Failure or delay in communicating lab or test results, assessment or consultant findings to the patient/caregiver	☐	☐
C. Failure to identify or address patient or caregiver concerns, preferences, or non-adherence	☐	☐

Please describe any significant "failure points" in relation to the patient's treatment/management. (e.g., patient was given wrong dose of __ medication, __ medication was delayed by 2 hours, medication wasn't available because of a shortage, patient did not receive PT because of lack of staff, etc.) You may also use this space to clarify any of your failure point choices, if applicable (e.g., if a failure point occurred multiple times, if you believe one failure point led to multiple diagnostic or treatment/management errors, etc.)	_____

If there was uncertainty about the diagnosis at discharge (e.g., no clear explanation for abdominal pain, altered mental status, etc.); OR 2) multiple diagnostic process failures were selected; OR 3) there was uncertainty about whether a diagnostic error may have occurred in this case, please review subsequent events (e.g., ambulatory visits, subspecialty visits, urgent care or ED visits, readmissions, major procedures or surgeries, expiration notes, autopsy findings), describe whether you think it could have been related to a missed opportunity to make an accurate and timely diagnosis during this hospital encounter.	_____
Please record here (in minutes) how long it took you to complete this chart review.	_____

APPENDIX B. Description of case, diagnoses, diagnostic error, harm, and process failures by risk cohort

ICU transfer 24 hours or more after admission		
Case: A 70+ year-old male with HTN, CKD, and HFrEF (40%) was admitted for evaluation of failure to thrive and weight loss. Upon arrival to the ED, his exam was notable for hypotension, 2+ lower extremity edema, and cool extremities. The NT-proBNP was elevated, troponins were rising, and lactate and creatinine were elevated. Intravenous fluids and antibiotics were initiated for presumed infection. Initial cardiology consultation suggested that hypotension was not cardiogenic, commenting "he is hypotensive though remains well-perfused". He was admitted to general medicine with a working diagnosis of sepsis and was continued on broad spectrum antibiotics. His blood cultures remained negative, and no clear infectious source was identified. Four days into hospitalization, he had a PEA arrest and was transferred to the ICU. TTE showed a severely depressed EF of 15%. He continued to deteriorate, requiring maximum inotropic and vasopressor support for mixed shock (cardiogenic +/- distributive), and expired.		
Diagnoses Primary Dx (Admission): Failure to thrive, weight-loss Primary Dx (Discharge): Mixed shock (cardiogenic +/- distributive) Secondary Dx: Hypotension	**Diagnostic Error:** Delay in diagnosing decompensated heart failure and cardiogenic shock upon admission despite compelling history, physical exam, and laboratory data. The primary team anchored on the initial cardiology consultant's assessment, placing too much weight on infectious and sub-optimally weighing a cardiac etiology. **Harm:** The patient experienced a cardiac arrest, was transferred to the ICU, continued to deteriorate, and ultimately expired. • Severity: Death • Definitely Preventable	**Diagnostic Domains & Specific Failures** • *Assessment:* failure or delay in considering the correct diagnosis; suboptimal weighing or prioritizing of primary or secondary diagnoses; too much weight to lower probability/priority diagnosis • *Access:* failure of triage and admission to wrong service • *Diagnostic Testing:* delay in ordering needed test(s); erroneous clinician interpretation of test • *Subspecialty Consultation:* suboptimal consultation or follow-up of consultation • *Other: History, Physical exam; Diagnostic Information and Patient Follow-up*
Death within 90 days of admission		
Case: A 40+ year-old female with biopsy-proven cirrhosis, optic glioma s/p resection, central hypothyroidism, anemia and thrombocytopenia, was admitted for evaluation of oliguric AKI. She underwent a thorough evaluation for decompensated cirrhosis and hepatorenal syndrome, including imaging, gastrointestinal and renal consultations, paracentesis, and endoscopy. She was placed on oxygen during the hospitalization. Towards the end of the hospital encounter, a CXR was obtained for concern of increased work-of-breathing by the night team. The official report commented on a small to moderate left pleural effusion, but this finding was never acknowledged by primary team members in their documentation. She remained on oxygen and was given a presumptive diagnosis of obstructive sleep apnea and obesity hypoventilation. Home oxygen was arranged. Three days after discharge, she was hospitalized for acute respiratory failure requiring intubation at which time she was diagnosed with left hepatic hydrothorax and pulmonary edema. She improved after a therapeutic thoracentesis. After subsequent hospitalizations related to end-stage cirrhosis, she was transitioned to hospice and expired peacefully at home.		
Diagnoses Primary Dx (Admission): Acute kidney injury, oliguria Primary Dx (Discharge): Acute kidney injury, oliguria Secondary Dx: Pleural effusion, hypoxia	**Diagnostic Error:** Missed opportunity to evaluate a unilateral pleural effusion in context of shortness of breath and worsening hypoxia. The overnight chest x-ray findings were not acknowledged in clinical documentation despite the final radiology report. **Harm:** The patient was readmitted within 24 hours at another hospital for pulmonary edema and hepatic hydrothorax which resolved after thoracentesis. • Severity: Major • Definitely Preventable	**Diagnostic Domains & Specific Failures** • *Assessment*: failure or delay in considering the correct diagnosis • *Diagnostic Testing*: failure or delay in ordering needed test(s); failure or delay in performing needed test(s); erroneous clinician interpretation of test • *Diagnostic Information and Patient Follow-Up*: failure or delay in acting on or following-up on test result (or results not communicated to the patient); missed physiologic monitoring finding (e.g., persistent hypoxia); failure or delay in recognizing or acting upon an urgent condition or complication • *Healthcare Team Communication and Collaboration:* failure or delay in communication between consultants and the patient's primary team
Complex clinical events including clinical deterioration (persistent fevers), acute kidney injury, multiple consultants		
Case: A 55+ year-old male with HLD presented to an outside hospital emergency room with pain in the right buttock and groin, was treated and discharged. He re-presented 2 days later with fever, chills, and acute urinary retention. He was referred to our ED for concern for spinal compression and potential intervention. In our ED, the MRI spine was unremarkable. On admission to general medicine, blood cultures returned positive for MSSA, however the source of bacteremia of remained uncertain. Infectious disease was consulted and initially suggested that the source was "from inoculation in a cold sore or through a blood draw or PIV placement when he first presented to the OSH". Fevers and bacteremia persisted despite appropriate, broad-spectrum antibiotics. A TEE was negative, and no other sources were pursued. The patient complained of persistent right buttock and groin pain for 4 days until repeat imaging was considered. A dedicated pelvic MRI obtained on hospital day 7 demonstrated right sacroiliac arthritis with multiple rim enhancing collections and edema in the surrounding musculature. The patient underwent wash-out by orthopedics. Subsequent hospital course was complicated by rash, eosinophilia, and AKI which was thought to be AIN from cefazolin per nephrology consultation. The patient was transitioned to daptomycin and discharged.		
Diagnoses Primary Dx (Admission): Sepsis	**Diagnostic Error:** Delay in diagnosing pelvic abscesses as source of bacteremia due to overweighing a lower probability source. There was a delay in obtaining dedicated pelvic MRI. Additionally, the infectious disease consultant did not consider alternative sources of persistent bacteremia despite appropriate antibiotic for several days.	**Diagnostic Domains & Specific Failures** • *Assessment:* failure or delay in considering the correct diagnosis; suboptimal weighing or prioritizing of primary or secondary diagnoses; too much weight to lower probability/priority diagnosis

Primary Dx (Discharge): Sepsis due to pelvic abscess Secondary Dx: MSSA bacteremia	**Harm:** The length of stay was prolonged and the hospital course was complicated. The patient was symptomatic for several days and required a major surgical intervention. The pelvic MRI showed evidence of complicated infection and destruction of surrounding tissues. The patient experienced complications related to his treatment. • Severity: Major Harm • Definitely Preventable	• *Diagnostic Testing*: failure or delay in ordering needed test(s); erroneous reading of test (lab/radiology); erroneous clinician interpretation of test • *Subspecialty Consultation:* suboptimal consultation or follow-up of consultation • *Other: History; Physical Exam; Diagnostic Information and Patient Follow-Up;*
None of the above criteria		
Case: A 70+ year-old female with Lewy-body dementia, prior right-sided MCA stroke, HTN and HLD was hospitalized for evaluation of AMS and a possible syncopal episode. She was hospitalized 1 month prior for cognitive decline, felt due to progressive dementia. Prior to this admission, her family reported that she lost consciousness for 1-3 minutes and her eyes were “rolling in the back of her head”. On presentation she was alert and oriented to name but had no focal deficits. Initial head CT and laboratory studies were unremarkable except for an abnormal urinalysis (pyuria, positive leukocyte esterase). On admission, the working diagnosis for AMS was an infectious etiology (UTI vs meningoencephalitis). The initial differential for syncope included vasovagal episode vs orthostatic episode vs arrythmia. Neurology was consulted and recommended routine EEG which was reported as having no epileptiform activity. She was treated for a UTI and discharged. The day after discharge the patient re-presented to the ED after the patient’s spouse expressed concerns about not being able to care for the patient at home given her mental status and lack of a clear diagnosis. A clinician documented "spoke with spouse who was verbalizing they sent patient home too early." Neurology re-evaluated the EEG from index hospitalization, felt it was more consistent with seizure activity and recommended anti-epileptic drug therapy.		
Diagnoses Primary Dx (Admission): Altered mental status Primary Dx (Discharge): Altered mental status Secondary Dx: Syncope, Lewy-body dementia	**Diagnostic Error:** Missed diagnosis of seizure as cause of acute change in mental status due to misinterpretation of EEG results at the time of discharge. The team under-weighed the possibility of seizures in the initial because of confounding of clinical features. **Harm:** The patient was readmitted within 24 hours and neurology confirmed seizure activity and started AEDs. • Severity: Moderate • Probably Preventable	**Diagnostic Domains & Specific Failures** • Assessment: suboptimal weighing or prioritizing of primary or secondary diagnoses • *History:* suboptimal weighing of a piece of history data • *Diagnostic Testing:* erroneous clinician interpretation of test; failure or delay in performing needed test • *Healthcare Team Communication and Collaboration*: failure or delay in communication between consultants and the patient’s primary team • *Patient Experience:* failure to communicate an accurate and timely explanation of the patient’s health problem(s) • *Other: Physical Exam; Subspecialty Consultation*