David E. Newman-Toker*, Zheyu Wang, Yuxin Zhu, Najlla Nassery, Ali S. Saber Tehrani, Adam C. Schaffer, Chihwen Winnie Yu-Moe, Gwendolyn D. Clemens, Mehdi Fanai and Dana Siegal*

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"

https://doi.org/10.1515/dx-2019-0104
Received December 8, 2019; accepted February 12, 2020

Abstract

Background: Missed vascular events, infections, and cancers account for ~75% of serious harms from diagnostic errors. Just 15 diseases from these "Big Three" categories account for nearly half of all serious misdiagnosis-related harms in malpractice claims. As part of a larger project estimating total US burden of serious misdiagnosis-related harms, we performed a focused literature review to measure diagnostic error and harm rates for these 15 conditions.

Methods: We searched PubMed, Google, and cited references. For errors, we selected high-quality, modern, US-based studies, if available, and best available evidence otherwise. For harms, we used literature-based estimates of the generic (disease-agnostic) rate of serious harms (morbidity/mortality) per diagnostic error and applied claims-based severity weights to construct disease-specific rates. Results were validated via expert review and comparison to prior literature that used different methods. We used Monte Carlo analysis to construct probabilistic plausible ranges (PPRs) around estimates.

Results: Rates for the 15 diseases were drawn from 28 published studies representing 91,755 patients. Diagnostic error (false negative) rates ranged from 2.2% (myocardial infarction) to 62.1% (spinal abscess), with a median of 13.6% [interquartile range (IQR) 9.2–24.7] and an aggregate mean of 9.7% (PPR 8.2–12.3). Serious misdiagnosis-related harm rates per incident disease case ranged from 1.2% (myocardial infarction) to 35.6% (spinal abscess), with a median of 5.5% (IQR 4.6–13.6) and an aggregate mean of 5.2% (PPR 4.5–6.7). Rates were considered face valid by domain experts and consistent with prior literature reports.

Conclusions: Diagnostic improvement initiatives should focus on dangerous conditions with higher diagnostic error and misdiagnosis-related harm rates.

Keywords: diagnosis; diagnostic errors; health services research; medical errors; misdiagnosis-related harms.

Introduction

Preventing medical misdiagnosis is a recognized national public health priority. In their landmark 2015 report *Improving Diagnosis in Healthcare*, the National Academy of Medicine (NAM) stated that "most people will experience at least one diagnostic error in their

***Corresponding authors: David E. Newman-Toker,** MD, PhD, Professor, Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA; Director, Armstrong Institute Center for Diagnostic Excellence, The Johns Hopkins University School of Medicine, Baltimore, MD, USA; and Professor, Department of Epidemiology, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, E-mail: toker@jhu.edu; and **Dana Siegal,** RN, CPHRM, CPPS, Director of Patient Safety, CRICO Strategies, 1325 Boylston Street, Boston, MA 02215, USA, Phone: +(617) 450-5552, Fax: +(617) 679-1237, E-mail: DSiegal@rmf.harvard.edu

Zheyu Wang and Yuxin Zhu: Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA; and Department of Biostatistics, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, E-mail: wangzy@jhu.edu (Z. Wang); daisy@jhu.edu (Y. Zhu)

Najlla Nassery: Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA, E-mail: nnasser3@jhmi.edu

Ali S. Saber Tehrani and Mehdi Fanai: Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA, E-mail: ali.tehrani@jhmi.edu (A.S. Saber Tehrani); mehdifanai@jhmi.edu (M. Fanai)

Adam C. Schaffer: Department of Patient Safety, CRICO, Boston, MA, USA; and Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, E-mail: ASchaffer@rmf.harvard.edu

Chihwen Winnie Yu-Moe: Department of Patient Safety, CRICO, Boston, MA, USA, E-mail: WYu@rmf.harvard.edu

Gwendolyn D. Clemens: Department of Biostatistics, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, E-mail: gclemen1@jhu.edu

Published online May 15, 2020

lifetime, sometimes with devastating consequences" [1]. Unfortunately, relatively little is known about the precise frequency of diagnostic errors and harms, leading the NAM to conclude that "the available research estimates [*are*] not adequate to extrapolate a specific estimate or range of the incidence of diagnostic errors in clinical practice today" [1].

Overall diagnostic error rates in real-world practice are not known, but a commonly cited estimate based on expert opinion is that 10–15% of all rendered diagnoses are incorrect [2]. Empiric rates measured in studies using chart reviews are often an order of magnitude lower. For example, a hospital-based study of 7926 records from hospital discharges and deaths found an error rate of 0.4% [3] and a primary care-based study of 1957 records estimated an error rate of 2.4% [4, 5]. This latter figure equates to more than 12 million Americans affected each year in primary care alone [5]. Assuming similar rates in emergency department and non-primary care ambulatory clinic visits, this would translate to ~31 million diagnostic errors annually. Such retrospective reviews are, however, limited by inadequate chart documentation [6, 7], low inter-rater reliability [8, 9], and hindsight bias [10]. A prospective study of 348 primary care visits using standardized patients found 13% of visits involved diagnostic errors [11], which is squarely in the range of expert opinion above. With ~1.3 billion US healthcare visits annually [12–14], the 10–15% estimate could translate to as many as ~100–200 million diagnostic errors each year.

Serious misdiagnosis-related harm rates are even less certain, but the same studies cited earlier found 0.22% of hospitalized patients [3] and 0.81% of primary care patients [5] suffered serious permanent morbidity or mortality from diagnostic error. With ~36 million hospitalizations annually [14], a 0.22% rate corresponds to ~80,000 serious harms (~40,000 deaths and ~40,000 disabilities) in US hospitals alone, which lands squarely in the ballpark of prior autopsy-based estimates suggesting 40,000–80,000 misdiagnosis-related deaths each year in US hospitals [15]. With ~0.5 billion US primary care visits annually [12], a 0.81% rate would translate to ~4 million serious harms in primary care alone [16]. With ~1.3 billion US healthcare visits annually [12–14], the same 0.22% and 0.81% rates would translate to ~10 million serious harms in total. Such large estimates for total serious harms, however, seem implausible – prior studies have found that roughly half of serious harms are deaths [17, 18], but there are far fewer than 5 million deaths per year in the US [19].

Diagnostic error and harm rates for specific diseases [2] or disciplines (e.g. radiology [20]) are often better quantified, but reported error rates still vary widely from ~1% to >50% [2] and harm rates vary from ~0.1% [21] to 45% [22] or more. Some of this variation is real, but much of it reflects differences in defining numerators and denominators for rates (Figure 1). The "numerator problem" (Figure 1, top) is that different standards for operationally defining a "diagnostic error" (or related concept) can substantially affect the rate. In a prospective study of 1152 patients with venous thromboembolism, short diagnostic delays (>1 week) occurred in 19.9%, while long delays (>3 weeks) occurred in only 4.9% [26]. If one counts any missed diagnostic finding or inter-rater variation as an error, rates can approach 100% [25]; if one counts only patients who return for additional care and are judged by chart review to have suffered preventable, misdiagnosis-attributable death, rates may be closer to ~0.01% [21]. Similarly, the "denominator problem" (Figure 1, bottom) is that the same number of measured errors or harms may be considered as a proportion of various target populations (e.g. all with the specific target disease vs. all encounters). For example, in radiology, error rates among radiographs with abnormal findings are roughly 30%, while error rates in typical radiologic practice (with a mix of predominantly normal and a few abnormal radiographs) are closer to 3.5–4% [20]. With rare but dangerous diseases such as spinal epidural abscess, the same number of serious harm events could be reported as 61% of misdiagnosed spinal abscesses, 34% of incident spinal abscess cases, or 0.0005% of all healthcare encounters [23]. These measurement and reporting differences often lead to confusion for readers, and they also make extrapolations to national estimates challenging.

Given the wide range of potential diagnostic error rates (~0.4% to ~100%), total diagnostic errors (~12 million to ~200 million per year in the US), serious misdiagnosis-related harm rates (~0.01% to ~45%), and total serious misdiagnosis-related harms (~40,000 to ~10 million), a new approach to estimation of diagnostic errors and harms is warranted. Across practice settings, missed vascular events, infections, and cancers (sometimes collectively referred to as the "Big Three" [16]) account for most of the morbidity and mortality attributable to diagnostic errors [27–30]. Using malpractice claims, we previously identified the five most frequent diseases in each "Big Three" category and showed that these 15 together account for nearly half of all serious harms [18]. As a second step toward a US national epidemiologic estimate of serious misdiagnosis-related harms, we sought to estimate diagnostic error and misdiagnosis-related harm rates for these 15 dangerous diseases using previously published medical literature.

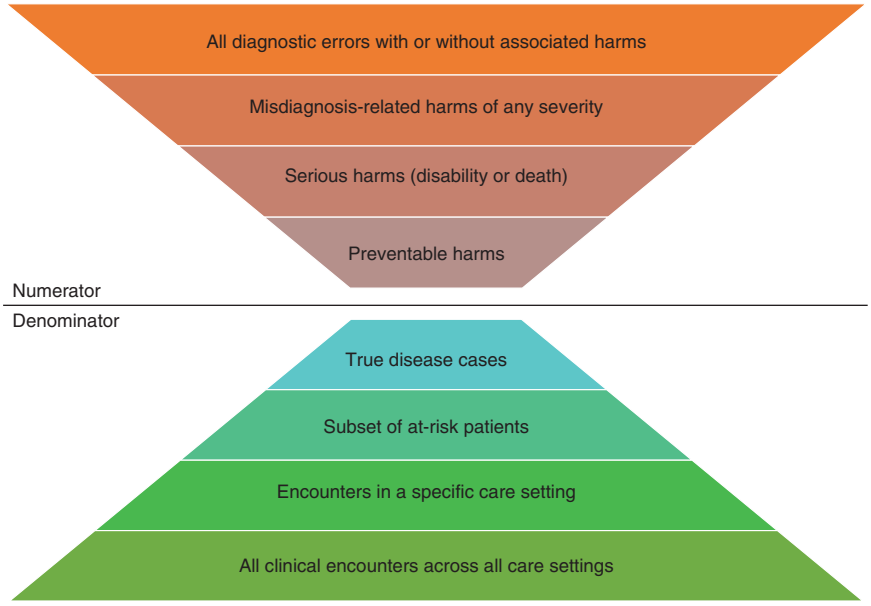

Figure 1: The "numerator and denominator problems" in comparing incidence of diagnostic errors across studies.
The top half of the figure shows potential numerators from larger (top) to smaller (bottom) that might be used to express diagnostic errors or harms. The bottom half of the figure shows potential denominators from smaller (top) to larger (bottom) that might be used to define the population in whom diagnostic errors or harms have occurred. With a large numerator (e.g. all diagnostic errors with or without harms) and a small denominator (e.g. all patients with a specific disease or subtype), measured error rates can exceed 50% for specific diseases in which errors are known to be frequent [23–25]. By contrast, with a small numerator (e.g. physician-judged preventable deaths) and a large denominator (e.g. all emergency department visits), we might expect a misdiagnosis-related harms incidence per visit closer to ~0.01% [21].

Materials and methods

Overall research concept

The overall goal of this three-phase research project was to construct a US national estimate of serious misdiagnosis-related harms (i.e. permanent disability or death). Each phase was designed to answer a key question from a specific data source that would support the final estimate: (1) what dangerous diseases account for the majority of serious misdiagnosis-related harms? (using malpractice claims data); (2) how frequent are diagnostic errors potentially causing harm among these dangerous diseases? (using medical literature-derived estimates from disease-specific studies); and (3) what is the overall epidemiologic incidence of diagnostic errors and harms among these dangerous diseases? (using nationally representative databases). The answer to the first question was recently published [18]. The answer to the second question is presented here. The answer to the third question will be presented elsewhere.

Current study design

We performed a focused review of the medical literature to estimate the rates of diagnostic errors and harms for the 15 "Big Three" diseases identified from the first phase of our three-phase project [18]. The list of 15 includes five key vascular events [i.e. stroke, myocardial infarction, venous thromboembolism, aortic aneurysm and dissection, and arterial thromboembolism (primarily acute mesenteric ischemia)]; five key infections (i.e. sepsis, meningitis and encephalitis, spinal abscess, pneumonia, and endocarditis); and five key cancers (i.e. lung cancer, breast cancer, colorectal cancer, prostate cancer, and melanoma). No human subjects were involved in this study, and no institutional review board (IRB) approval was required.

We sought to identify the highest-quality, modern, US-based studies, if available, and, otherwise, the best available evidence if the prior criteria could not be met. Because few disease-specific studies on misdiagnosis also assessed misdiagnosis-related harms, we used literature-based estimates of the overall, generic (disease-agnostic) serious misdiagnosis-related harm rates per diagnostic error and combined them with data from malpractice claims and population incidence to construct disease-specific rates. An overview of the steps taken to derive the final rate estimates is provided in Figure 2.

Although more accurately called *proportions*, here we use the more common term, *rates*. We only included diagnostic false-negative rates (1 – sensitivity) [31] in our literature search because the third phase of our three-phase project relies on population-based incidence data for the 15 specific diseases. The incidence data are based on presumed "true disease" (i.e. inpatient hospital discharge diagnoses for vascular events and infections and cancer registry data for cancers). The population incidence of false-positive diagnoses can only be calculated from data on the incidence of "not true disease", which cannot be measured unless data on clinical presenting symptoms and cohort follow-up are available [32].

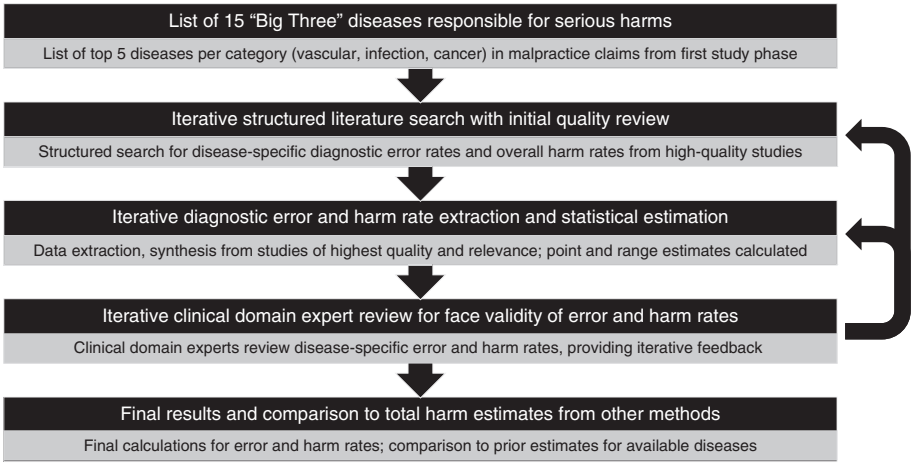

Figure 2: Overview of methods for literature-derived diagnostic error and harm rate estimation.
Looping back arrows on the right side of the flow chart refer to iterative steps sparked by feedback from clinical domain experts.

Definitions

Published definitions were used for diagnostic error and misdiagnosis-related harms. The NAM defines a *diagnostic error* 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]. *Misdiagnosis-related harms* are defined as harms resulting from the delay or failure to treat a condition actually present, when the working diagnosis was wrong or unknown [delayed or missed diagnosis (false negative)], or from treatment provided for a condition not actually present [wrong diagnosis (false positive)] [33, 34]. For the purposes of our search, harms from false positives were not considered (see Current study design). Harm severity was defined according to the National Association of Insurance Commissioners (NAIC) Severity of Injury Scale, a recognized insurance industry standard for measuring severity of injury in malpractice claims [35, 36]. *Serious (high-severity) misdiagnosis-related harms* were defined using NAIC severity codes 6–9, representing serious permanent morbidity (NAIC 6–8) or mortality (NAIC 9) [18].

Literature search strategy

Given the need to assess error and harm rates for a large number of diseases, we did not attempt to conduct a full systematic review and meta-analysis for each disease. Nevertheless, literature searches were conducted by authors with substantial prior experience in systematic reviews and meta-analysis (A.S.T., D.N.T.). We searched PubMed, Google Scholar, Google, and our personal files for relevant citations. We began with an initial PubMed search using a standardized search string focused on diagnostic errors (Supplementary Material A1). We cross-referenced this search with terms related to each of the 15 diseases. When a large number of citations were encountered, we further cross-referenced the search with terms related to meta-analysis. When few results returned, we pursued additional searches using broader diagnostic error-related searches (e.g. not restricted to title words). We also searched using PubMed tools (e.g. "Similar Articles" search), the reference lists of identified references, and the reference lists of prior narrative reviews of diagnostic error rates [2]. Some error rates were harder to identify. In particular, most of the high-quality literature we found for cancer misdiagnosis was eventually identified using terms describing diagnostic "intervals" rather than "delays".

Diagnostic error and harm rates

Diagnostic error rates were abstracted from the highest quality literature we could identify. We discarded lower-quality studies when more rigorous studies (e.g. systematic reviews, population-based sampling, large sample sizes, rigorous case ascertainment) were available. We quantitatively combined studies of comparable quality for point estimates or ranges, as appropriate (see Statistical analysis and reporting). Because hospital autopsy-detected diagnostic error rates have been declining over time [29, 37], we sought to identify temporal trends in diagnostic error rates whenever possible.

Because most disease-specific diagnostic error studies did not measure frequency or severity of misdiagnosis-related harms, we were forced to use a generic per-error harm rate and then construct a disease-specific weight using additional information from malpractice claims. To estimate the generic (disease-agnostic) rate of serious harms per diagnostic error, we took data from the five large, well-respected studies of diagnostic errors from general care settings that identified both harm frequency and severity [3, 4, 38–40]. To estimate disease-specific harm rates, we used the proportion of high-severity harms from diagnostic error malpractice claims cases from the first phase of this project [18] to construct a disease-specific harm-severity weighting. The diagnostic error rate was multiplied by the weighted disease-specific harm rate to get a disease-specific, serious misdiagnosis-related harm rate per incident case of disease (i.e. the combined error-harm rate for each of the 15 diseases). Using additional information available from the first phase of our project [18], we also constructed estimates for "other" vascular events, infections, and cancers. Additional details may be found in Section A2 of the Supplementary Material.

As a further check on our diagnostic error and serious harm rates, we sent the estimates to 25 relevant domain experts and asked for their feedback regarding face validity of the error and harm rates in their respective areas of domain expertise. During that process, we also made note of additional feedback provided by experts on the framing and interpretation of our results. We also compared our final

serious misdiagnosis-related harm rates to empiric rates found in rigorously designed studies, where available.

Statistical analysis and reporting

We report point estimates for error and harm rates with either 95% confidence intervals (CIs) or plausible ranges (PRs), as appropriate. If available, we used 95% CIs reported in source manuscripts. When homogeneous studies of comparable quality were available, we combined results to determine a mean error rate and then calculated upper and lower 95% CI bounds. When studies used more than one threshold cutoff for defining a missed or delayed diagnosis (e.g. based on short vs. long diagnostic delay) or revealed heterogeneous results, we used input from domain experts to help choose the most appropriate point estimate or PR bound. For combined error-harm rates (i.e. the arithmetic product of disease-specific diagnostic error and per-error harm rates), we derived variability estimates using a probabilistic sampling approach based on Monte Carlo simulations [41] (Supplementary Material A3). In this paper, these ranges are denoted as "probabilistic plausible ranges" (PPRs), rather than 95% CIs, to caution against the fact that some diagnostic error rates use PRs rather than 95% CIs as their range, reflecting uncertainty beyond mere sampling error. We calculated 95% CIs around "rates" (i.e. proportions) using Stata v14.2 (College Station, TX, USA) and PPRs using R v3.4.4 (Vienna, Austria). This paper follows EQUATOR (STROBE) [42] reporting guidelines for observational studies.

Results

Diagnostic error rates

Condition-specific diagnostic error rates were derived from high-quality meta-analyses, large prospective clinical trials, or studies using population-based sampling for 14 of 15 diseases (Tables 1–3). For arterial thromboembolism, the best available evidence was summarized from four retrospective, single-center studies (Table 1). Drawn from 28 studies collectively representing 91,755 patients, the total per-disease individual study sample sizes were greatest for cancers (mean 14,690; median 11,860), intermediate for vascular events (mean 3068; median 1532), and least for infections (mean 593, median 309). We observed that specific operational definitions for diagnostic error influenced their estimated frequency substantially. For example, diagnostic error for venous thromboembolism occurred in 19.9% when using a delay cutoff of >1 week but just 4.9% when using a delay of >3 weeks [26].

Disease-specific diagnostic error (false negative) rates ranged from 2.2% (myocardial infarction) to 62.1% (spinal abscess), with a median disease-specific rate of 13.6% [interquartile range (IQR) 9.2–24.7] across the 15 individual diseases. The aggregate mean rate across all of these 15 "Big Three" diseases was 9.7% (8.2–12.3). Studies that assessed diagnostic error over an extended time period and analyzed for trends in misdiagnosis rates found no decline (stroke, 1996–2016 [31]) or small declines that were not statistically significant (aortic aneurysm, 1961–2005 [44]; aortic dissection, 1996–2007 [45]). For colorectal cancer, rates of diagnostic delay were lower in a study from 2010 to 2014 [62] than in a study from 1998 to 2005 [63] (Table 3); experts felt that increases in colorectal cancer screening over time could have been partly responsible, but other methodological differences between the two studies could also have explained the discrepancy.

Misdiagnosis-related harm rates

The generic rate of serious misdiagnosis-related harms, derived from five major studies of diagnostic error across care settings (Table 4), was 30.8% (n = 374/1216, 95% CI 28.2%–33.4%) [3, 4, 38–40]. Schiff et al. described 583 physician-reported diagnostic errors based on surveys during the years 2002–2004 from a convenience sample of 283 physicians from 22 institutions in six states that included a mix of general internists, medical specialists, and emergency physicians, 47% of whom identified themselves as primary care physicians [38]. They identified case severity as "major" in 28% (defined as "death, permanent disability, or near life-threatening event"); there was no difference in the rate of a case being considered "major" whether physicians were reporting their own diagnostic error vs. someone else's diagnostic error [38]. Zwaan et al. examined 7926 hospital discharges and deaths from a stratified, random sample of Dutch hospitals (n = 40) in 2004 for adverse events, defined as "(1) an unintended (physical or mental) injury that (2) resulted in prolongation of the hospital stay, temporary or permanent disability, or death and (3) was caused by health care management rather than the patient's disease" [3]. They found 80 diagnostic adverse events, 29% of which resulted in death and 26% of which resulted in disability at discharge [3]. Singh et al. reviewed records from 1957 primary care visits to 69 primary care providers in 2006–2007 at two large health systems identified using electronic health record trigger tools [4]. They identified 190 instances of diagnostic error, 14% of which were judged to have resulted in "immediate or inevitable death" and 19% of which resulted in "serious permanent damage" [4]. Ely et al. described 202 physician-reported diagnostic errors based on surveys during the years 2009–2010 from a stratified random sample of 200 family physicians, 200 general internists, and 200 general pediatricians practicing in Iowa and registered with the Iowa Board of Medical Examiners. Among 184 errors where outcomes were known, 27% resulted in death and 13% resulted in

Table 1: Literature-derived missed or delayed diagnosis rates for the top 5 vascular events.

Vascular event	Point estimate[a]	Lower bound	Upper bound	Study design (sample)	Notes
Stroke	8.7%	8.0% (95% CB)	9.3% (95% CB)	Meta-analysis (23 studies; n = 10,536 patients) [31]	23 studies (12 US-based) from 1995 to 2016; error rates varied by clinical presentation, with milder, non-specific, transient symptoms having the highest error rates (range 24–60%, OR 7–14)
Myocardial infarction	2.2%	1.6% (95% CB)	3.0% (95% CB)	Prospective clinical trial (n = 1855) [43]	10,689 suspected acute coronary syndromes in 10 US EDs in 1993; erroneously discharged myocardial infarction 2.1% (n = 19/889) or unstable angina 2.3% (n = 22/966)
Venous thromboembolism	19.9%	4.9% (PB % with long delay[b])	22.3% (95% CB)	Prospective observational (n = 1152 patients) [26]	1152 hospitalizations at 70 medical centers (68 in US) in 1999; mix of delays in seeking care (esp. DVT) and after medical attention (PE > DVT); 19.9% (n = 229) with delays (>1 week); 4.9% (n = 57) with very long delays (>3 weeks), same in both groups (PE, DVT)
Aortic aneurysm and dissection	27.9%	25.6% (95% CB)	30.2% (95% CB)	(1) Meta-analysis [nine studies; n = 1109 patients (four studies since 1990, n = 638[c])] [44] (2) International registry [n = 894 (51.7% of cases since 2000)] [45]	(1) Ruptured abdominal aneurysm error rate 42% overall (95% CI 29–55%) and 32% (95% CI 16–49%) in more recent studies reported during 1990–2006 (three Europe, one New Zealand); error rate difference in studies reported before vs. after 1990 was not statistically significant (2) International registry (24 centers, 11 countries) of acute aortic dissections; analyzed thoracic dissections 1996–2007 (n = 1204, 894 with time data available for analysis); time to diagnosis >4 h in 50%, ≥24 h in 25%; 4.5 h before 2000, 4.2 h after 2000, difference not statistically significant; North American centers (disproportionately located in the US) had longer delays than European centers
Arterial thromboembolism	23.9%	18.9% (95% CB)	29.5% (95% CB)	Retrospective chart reviews (total n = 264 patients across four unrelated studies) [46–49]	(1) Delay >24 h in 25.0% (n = 15/60) with acute mesenteric ischemia from embolism (n = 20) or thrombosis (n = 40) (university hospital, Germany, 2000–2006) (2) Delay >24 h to surgical consult in 15.3% (n = 11/72) (university hospital, US, 2004–2005) (3) Mesenteric ischemia radiographic findings missed in 34.3% (n = 12/35) on CTA in consecutive acute abdomen patients suspected of mesenteric ischemia clinically and confirmed at surgery (university hospital, Italy, 2007–2011) (4) Delay >5 h in ~50%, >24 h in ~26% of 97 with acute mesenteric ischemia from mixed causes (tertiary care center, Finland, 2009–2013)

[a]Diagnostic error rates reported are false-negative rates (i.e. missed/delayed). Studies generally defined these rates based on one of two strategies: (1) encounters at which the diagnosis might have been made, but was not (i.e. missed opportunities); (2) absolute diagnostic delay relative to the urgency of illness detection, as defined by disease natural history. In the latter case, the time window to avoid harm was necessarily disease-specific and, therefore, defined differently across studies of different diseases (e.g. hours for aortic dissection, months for cancer). [b]When shorter and longer delays were reported, we chose either the longer delay (more conservative) or the shorter delay (less conservative) for the point estimate, based on feedback from experts. We then used the other estimate to define the upper or lower plausible bound, rather than the 95% CI bound (i.e. widening the estimated plausible range beyond the 95% CI for this condition). When results were heterogeneous across included studies, we used different studies to define the point estimate and one or both bounds for the plausible range. [c]Given the evolution in routine use of abdominal CT scans for definitive diagnosis after 1990 (and lower error rates thereafter), we considered only the four modern studies (published during 1998–2006) from Azhar et al.'s meta-analysis of ruptured aortic aneurysms when calculating our point estimate. CB, confidence bound; CT, computed tomography; CTA, computed tomography angiography; DVT, deep vein thrombosis; ED, emergency department; OR, odds ratio; PB, plausible bound; PE, pulmonary embolus.

Table 2: Literature-derived missed or delayed diagnosis rates for the top 5 infections.

Infection	Point estimate[a]	Lower bound	Upper bound	Study design (sample)	Notes
Sepsis	9.5%	8.2% (PB[b])	20.8% (PB[b])	(1) Population-based retrospective cohort (n = 332 sepsis) [50] (2) Regional health network retrospective cohort (n = 1094 sepsis) [51] (3) Ambispective ED-based cohort (n = 110 infection, n = 54 sepsis) [52] (4) Retrospective hospital-based cohort with systematic sampling (n = 300 sepsis) [53]	(1) Pediatric hospitalizations with sepsis, Canada 2005–2010, 20.8% (n = 69/332) initially missed/discharged from ED; drawn from a 5-year sample of 2.4M ED visits for any symptom or problem in the age cohort (2) Pediatric hospitalizations with sepsis, US 2014–2015, 2.9% (n = 32/1094) delayed diagnosis during hospitalization; drawn from a 2-year sample of 280,884 visits; large disparity noted between academic false-negative rate (2.0%, n = 20/996) and community false-negative rate (12.2%, n = 12/98) (3) Consecutive adult ED visits (n = 487) to a university hospital in Germany, 2013; 110 found to have infections on hospitalization, 54 with sepsis; sepsis not initially recognized in ED in 59.3% (n = 32/54) (4) Randomly selected adult, in-hospital deaths, or hospice discharges (n = 568) 2014–2015 from academic (n = 3) and community (n = 3) hospitals; reviewed and identified 300 sepsis cases, 198 sepsis deaths; of the 300, 36 were considered definitely (n = 4), probably (n = 7), or possibly (n = 25) preventable, had earlier diagnosis led to prompt administration of antibiotics
Meningitis and Encephalitis	25.6%	20.8% (95% CB)	30.8% (95% CB)	Population-based retrospective cohorts (two unrelated studies, total n = 309 meningitis) [50, 54]	(1) Pediatric hospitalizations with meningitis, Canada 2005–2010, 23.8% (n = 45/189) initially missed in ED (2) Pediatric hospitalizations with meningitis, Australia 1994–1999, 28.3% (n = 34/120) initially missed in ED
Spinal abscess	62.1%	54.6% (95% CB)	69.2% (95% CB)	(1) Retrospective chart review with systematic case ascertainment (sampled n = 119) [23] (2) Retrospective case-control chart review with systematic case ascertainment (n = 63) [22]	(1) Large database from Department of Veterans Affairs (VA) (1700 facilities, 8 million US patients); 446 new spinal abscesses in 2013; randomly selected 250 charts for dual independent review; excluded 131 diagnosed outside the VA; delayed diagnosis determined in 55.5% (n = 66/119) (2) All patients discharged with a diagnosis of spinal abscess, vertebral osteomyelitis, or discitis screened at a single, urban, US-based academic ED 1992–2002; abscess cases all confirmed by MRI, CT, or operative reports, 74.6% (n = 47/63) had delay [22]
Pneumonia	9.5%	2.3% (PB[c])	14.3% (95% CB)	(1) Prospective, multicenter, cross-sectional study of ED patients with suspected community-acquired pneumonia (n = 319), 163 of these confirmed [55]; (2) Prospective symptom cohort of consecutive patients with dyspnea and 95% follow-up (n = 247 dyspnea, 47 with pneumonia) [56]	(1) Consecutive patients with lower respiratory tract symptoms at four academic EDs in France 2011–2013; 8.6% (n = 14/163) of those clinically deemed unlikely to have pneumonia were confirmed to have pneumonia (2) 247 patients with dyspnea at five acute care hospitals in the Netherlands 2007–2008 (47 with pneumonia); 13.8% overall misdiagnosis rate (same error rate for pneumonia, per lead author, Laura Zwaan, personal communication)

Table 2 (continued)

Infection	Point estimate[a]	Lower bound	Upper bound	Study design (sample)	Notes
Endocarditis	25.5%	21.7% (95% CB)	29.6% (95% CB)	Prospective study with population-based sampling (n = 486) [57]	486 patients from seven regions in France with definite infective endocarditis in 2008; late diagnosis (>1 month after onset of first symptoms) occurred in 25.5% (n = 124/486)

[a]Diagnostic error rates reported are false-negative rates (i.e. missed/delayed). Studies generally defined these rates based on one of two strategies: (1) encounters at which the diagnosis might have been made, but was not (i.e. missed opportunities); (2) absolute diagnostic delay relative to the urgency of illness detection, as defined by disease natural history. In the latter case, the time window to avoid harm was necessarily disease-specific and, therefore, defined differently across studies of different diseases (e.g. hours for aortic dissection, months for cancer). [b]The estimates across all four studies for missed sepsis were heterogeneous. In particular, from two studies that included both academic and community hospitals, the reported academic hospital rate of missed sepsis (2.7%, 95% CI 1.9–3.8) was far lower than the community hospital rate (17.5%, 95% CI 14.5–20.8). Accordingly, we assigned a wider plausible bound to the range. Specifically, we used the lower 95% CI bound of the total sample (9.5%, 95% CI 8.2–11.0), and the upper 95% CI bound of the community-only subsample from the same four studies (20.8%). [c]Claessens et al. [55] identified 14 patients clinically deemed unlikely to have pneumonia who were confirmed to have pneumonia. Among these, four of 14 (n = 4/163) were still deemed unlikely by clinicians even after a chest CT scan result. To define the lower plausible bound, we combined the lower rate from Claessens et al. with the results from the study by Zwaan et al. [56] (4.8%, 95% CI 2.3–8.6) and then used the lower 95% CI bound. CB, confidence bound; ED, emergency department; PB, plausible bound; VA, Veterans Affairs.

"permanent disabilities" [39]. Okafor et al. evaluated 509 voluntary incident reports at two large academic-affiliated emergency departments from 2009 to 2013. They identified 209 diagnosis-related incidents, 16% resulting in "major harm" defined as a "life-threatening or limb-threatening event, permanent disability or death" [40]. Where the causes of harm were described, the harms were disproportionately due to illness progression, rather than adverse consequences of treatments for other (e.g. benign) diseases that the patient did not actually have (i.e. wrong diagnoses).

Disease-specific harm severity weights, severity-weighted serious misdiagnosis-related harm rates per error, and serious misdiagnosis-related harm rates per incident disease case are shown in Table 5. The disease-specific rates of serious harms per incident case of disease ranged from 1.2% (myocardial infarction) to 35.6% (spinal abscess), with a median disease-specific rate of 5.5% (IQR 4.6–13.6) across the 15 individual diseases. The aggregate mean rate across all of these 15 "Big Three" diseases was 5.2% (PPR 4.5–6.7). Put differently, we estimated that one of every 85 patients with a myocardial infarction, roughly one of every 20 patients with any top 5 "Big Three" disease, and more than one of every three patients with a spinal abscess suffers death or permanent disability as a consequence of being misdiagnosed.

Expert validation of error and harm rates

To assess the face validity of our literature-derived estimates for diagnostic error and serious harm rates, we sent our results to 25 domain experts not part of the study and asked for their impression as to the plausibility of the estimates in their area of expertise. We received feedback from 23 physicians, including individuals trained in cardiology (x1); dermatology (x1); emergency medicine (x6); gastroenterology (x1); infectious diseases (x3); neurology (x2); oncology (x5); radiation oncology (x1); and surgery (x3). All but one agreed to be recognized for their input (see Acknowledgments). After the first set of exchanges with experts, we adjusted several of our estimates (in each of the "Big Three" categories), before returning our revised estimates to them for final review and approval.

For vascular events, our initial estimate for missed myocardial infarction (0.8%, based on a large administrative data study [69] that we ultimately did not use for estimation) was deemed too low by experts from emergency medicine and cardiology. They felt that the estimate from the older but more robust patient-level clinical trial dataset [43] (2.2%, upper bound 3.0%) was more realistic. We switched to exclusively using the latter study for

Table 3: Literature-derived missed or delayed diagnosis rates for the top 5 cancers.

Cancer	Point estimate[a]	Lower bound	Upper bound	Study design (sample)	Notes
Lung cancer	22.5% [58]	11.3% [59] (PB % with long delay[b])	37.8% [60] (PB % with any delay[b])	(1) National registry study (n = 32,441) [58] (2) Interview study (two states, five centers or clinics; n = 275) [59] (3) Retrospective cohort study (2 medical centers; n = 587) [60]	(1) NCI SEER Database 2003–2006 linked to Medicare; 12 registries representing nine states, 14% of US population; guideline-discordant delays 22.5% (n = 7302), almost entirely due to diagnostic delay (2) Two-state, five cancer centers/clinics, direct patient interviews; 11.3% (n = ~31/275) perceived diagnostic delay and took >90 days to treatment (3) Two Department of Veterans Affairs (VA) medical centers in the US; 587 new pathologically diagnosed lung cancers 2004–2007; dual independent review of charts; 222 judged to have ≥1 missed diagnostic opportunity
Breast cancer	8.9%	8.5% (95% CB)	26.3% (PB % with short delay[b])	National registry study (n = 21,818) [61]	National Cancer Comprehensive Network (NCCN) Breast Cancer Outcomes Database 2000–2007; eight US comprehensive cancer centers; delay >60 days in 26.3% (n = 5747) and >180 days in 8.9% (n = 1937)
Colorectal cancer	9.6% [62]	8.4% [62] (95% CB)	47.7% [63] (PB % with short delay[b])	(1) National registry study (n = 10,663) [63] (2) Retrospective cohort study (single, large health plan/system; n = 2191) [62]	(1) NCI SEER Database 1998–2005 linked to Medicare; 12 registries representing nine states, 14% of US population; delay ≥60 days in 47.7% (n = 4614/9669)[c] (2) Two divisions of a single large health plan/system (Kaiser Permanente Northern and Southern California); 50–70 year olds, 2010–2014 (n = 70,124), with a positive screening fecal immunochemical test for colorectal cancer followed by diagnostic colonoscopy; 2191 found to have colorectal cancer; delay ≥10 months in 9.6% (n = 211)
Prostate cancer	2.4%	1.7% (95% CB)	13.8% (PB % with short delay[b])	National registry study (n = 1763) [64]	UK Clinical Practice Research Datalink 1998–2009; 600 primary care practices, 7% of UK population; delay despite red flag symptoms >1 month in 13.8% (n = 244) and >6 months in 2.4% (n = 42)
Melanoma	13.6%[d]	6.8% [65] (PB % with long delay[b] after told "all clear")	25.0% [66] (PB % with short delay[b])	(1) Population-based interview study using registry data (n = 3772) [65, 66] (2) Retrospective cross-sectional chart review study (n = 933)[d]	(1) Population-based telephone survey in Queensland, Australia for patients diagnosed with invasive melanoma 2000–2003 in the Queensland Cancer Registry; post-presentation (first physician to definitive diagnosis) delay times of >6 weeks in 25% [66]; delay >3 months after first physician said either "watch it for a while" (6.9%) or "all clear" (6.8%) in a combined 13.7% [65] (2) Cross-sectional study (2009–2015) at a single, US-based academic center; 933 pathologically proven melanomas (591 pigmented, 342 amelanotic); chart review ~156 clinically misdiagnosed, ~82 pathologically misdiagnosed[d]

[a]Diagnostic error rates reported are false-negative rates (i.e. missed/delayed). Studies generally defined these rates based on one of two strategies: (1) encounters at which the diagnosis might have been made, but was not (i.e. missed opportunities); (2) absolute diagnostic delay relative to the urgency of illness detection, as defined by disease natural history. In the latter case, the time window to avoid harm was necessarily disease-specific and, therefore, defined differently across studies of different diseases (e.g. hours for aortic dissection, months for cancer). [b]When shorter and longer delays were reported, we chose either the longer delay (more conservative) or the shorter delay (less conservative) for the point estimate, based on feedback from experts. We then used the other estimate to define the upper or lower plausible bound, rather than the 95% CI bound (i.e. widening the estimated plausible range beyond the 95% CI for this condition). When results were heterogeneous across included studies, we used different studies to define the point estimate and one or both bounds for the plausible range. [c]Denominator for short-delay proportions was based on summed denominators from table 4 of Pruitt et al. Although not described in the paper, this total (n = 9669) is less than the total study sample (n = 10,663) [63], presumably because of missing or censored data. [d]We included Strazzulla et al. [67], despite its weaker study design than Baade [65, 66], because the study reflected more recent US-based data and had very similar error rates and ranges to those from the stronger, population-based data from Australia. We normalized data from Strazzulla to the population-based relative prevalence of pigmented (92.1%) vs. amelanotic (7.9%) melanomas using Thomas et al. [68] We did this because the Strazzulla data came from a single referral center and over-represented amelanotic melanomas relative to the general population (36.7% vs. 7.9%), which are misdiagnosed at higher rates. The consequence of the normalization was to reduce the diagnostic error rate estimate from 16.8% to 13.0%. A mean was then calculated combining this value with 13.7% from Baade et al., 2007 for the point estimate of 13.6%. CB, confidence bound; NCI, National Cancer Institute; PB, plausible bound; PSA, prostate-specific antigen; SEER, Surveillance, Epidemiology, and End Results program/database; VA, Veterans Affairs.

Table 4: Generic (disease-agnostic) rates of serious misdiagnosis-related harms per diagnostic error.

Source	Study method	Population	# Diagnostic errors	# Serious harms[a]	% Serious harms
Schiff et al. [38] (2002–2004)	Physician report (survey)	Mixed, 47% primary care	553	159	28.8%
Zwaan et al. [3] (2004)	Record review (triggered)	Inpatient	80	44	55.0%
Singh et al. [4] (2006–2007)	Record review (triggered)	Primary care	190	63	33.2%
Ely et al. [39] (2009–2010)	Physician report (survey)	Primary care	184[b]	74	40.2%
Okafor et al. [40] (2009–2013)	Physician report (incidents)	Emergency department	209	34	16.3%
Total	–	–	1216	374	30.8%

[a]Serious harm was named and defined slightly differently in each study (see Results), but essentially reflected either death or permanent disability as a health outcome state, which is very similar to our definition in the current study. [b]Ely et al. described 202 diagnostic errors in their article, but they had no outcome information on 18 patients. Thus, the % serious harm is calculated out of 184 here.

the point estimate and associated CI. This was the only error or harm estimate revised *upward* through the expert review process. One expert questioned whether the misdiagnosis rate for arterial thromboembolism, based on weaker sources, might be too high, but the others did not, and we left this unchanged. Several experts mentioned they would have preferred to separate aortic aneurysm rupture from aortic dissection, as their clinical presentations differ; however, these were left grouped together because error and harm rates are similar, and these are grouped together in the standard classification schema that we used to define all the disease groupings [18]. For infections, our initial estimates for harm from sepsis and pneumonia misdiagnosis were deemed too high; additional literature was identified for sepsis, and an outlier study for pneumonia misdiagnosis was removed.

Our initial estimates of delayed cancer diagnosis rates were almost uniformly considered too high by our experts. We had originally chosen as the point estimate the shorter diagnostic delay period (i.e. higher diagnostic error rate) reported in each of the studies and used the longer delay period (i.e. lower diagnostic error rate) as the lower PR bound. After expert feedback, we switched the point estimates from the higher rates to the lower rates – the cancer experts then found the revised estimates face valid. It was clear that diagnostic error rates and per-error harm rates could not be completely disaggregated in experts' minds, particularly for cancers. In other words, the threshold for considering a delay a "diagnostic error" (i.e. missed or delayed diagnosis) was inextricably linked to the question of its impact – short delays might cause zero harm, while long delays might be associated with frequent, severe harms (e.g. death). Several of our experts were therefore much more comfortable judging the harm rate per incident disease case (i.e. the mathematical product of diagnostic error rate and per-error harm rate), rather than these two numbers separately. The final serious harm rates per incident cancer case [range 1.2–13.8% for five specific cancers (Table 5)] were judged believable by experts. This review process thus accounts for the anticipated inverse relationship between error and per-error harm rates (Supplementary Material B1).

Finally, five of our six emergency physicians expressed specific concerns related to their practice context when measuring and reporting diagnostic errors and misdiagnosis-related harms. Three key issues were raised in their comments: (1) emphasizing that diagnostic *delay* (not only "wrong" or "missed") is part of the "diagnostic error" rate estimates, as articulated in the NAM definition used here ("accurate and timely") [1]; (2) clarifying that such errors do not only occur in the emergency department (e.g. aneurysmal subarachnoid hemorrhage, where misses of this acute disease disproportionately occur in primary care [70]); and (3) acknowledging that 100% diagnostic accuracy is unattainable and the pursuit of 100% accuracy could have adverse, unintended consequences that cause more harm than good (e.g. harms from testing, management of incidental findings, or overdiagnosis leading to overtreatment). All agreed, however, that there remain opportunities to improve on current diagnostic performance in the emergency department.

Literature validation of error and harm rates

As a final check on the validity of our results, we compared our misdiagnosis-related harm rates to previously published estimates using other methods. At first glance, the per-error harm rates (Table 5) seemed to us to be quite high. However, these estimates were similar

Table 5: Severity-weighted, disease-specific diagnostic error and serious misdiagnosis-related harm rates.

Big Three disease	Diagnostic error rate % (95% CI, PR, PPR[a])	Disease-specific harm severity weight	Severity-weighted serious harm rate per diagnostic error (PPR)	Serious misdiagnosis-related harm rate per incident case of disease (PPR)
Vascular				
Aortic aneurysm and dissection	27.9% (CI: 25.6–30.2)	1.98	60.9% (55.9–66.1)	17.0% (15.1–19.1)
Arterial thromboembolism	23.9% (CI: 18.9–29.5)	1.70	52.3% (48.0–56.8)	12.5% (9.7–15.7)
Venous thromboembolism	19.9% (PR: 4.9–22.3)	1.70	52.3% (48.0–56.8)	10.4% (2.6–11.8)
Stroke	8.7% (CI: 8.0–9.3)	1.80	55.2% (50.7–60.0)	4.8% (4.3–5.3)
Myocardial infarction	2.2% (CI: 1.6–3.0)	1.73	53.2% (48.9–57.8)	1.2% (0.8–1.6)
Top 5 vascular events subtotal	8.7% (PPR: 6.8–9.1)	1.77	54.4% (52.4–57.0)	4.7% (3.7–5.0)
Other vascular events	8.7% (PPR: 6.8–9.1)[b]	0.39	11.9% (10.9–12.9)	1.0% (0.8–1.1)
Total vascular events	8.7% (PPR: 6.8–9.1)	1.03	31.7% (30.7–33.1)	2.8% (2.2–2.9)
Infection				
Spinal abscess	62.1% (CI: 54.6–69.2)	1.88	57.9% (53.2–62.9)	36.0% (30.9–41.3)
Meningitis and encephalitis	25.6% (CI: 20.8–30.8)	1.83	56.3% (51.7–61.2)	14.4% (11.5–17.7)
Endocarditis	25.5% (CI: 21.7–29.6)	1.72	52.9% (48.6–57.5)	13.5% (11.3–16.0)
Sepsis	9.5% (PR: 8.2–20.8)	1.88	57.9% (53.1–62.9)	5.5% (4.7–12.1)
Pneumonia	9.5% (CI: 2.3–14.3)	1.55	47.8% (43.8–51.9)	4.5% (1.1–6.9)
Top 5 infections subtotal	10.2% (PPR: 6.9–15.4)	1.72	52.9% (50.8–58.4)	5.4% (3.8–8.4)
Other infections	10.2% (PPR: 6.9–15.4)[b]	0.99	30.6% (28.1–33.2)	3.1% (2.1–4.8)
Total infections	10.2% (PPR: 6.9–15.4)	1.34	41.1% (39.6–44.1)	4.2% (2.9–6.4)
Cancer				
Lung cancer	22.5% (PR: 11.4–37.8)	2.01	61.9% (56.8–67.2)	13.9% (7.0–23.6)
Melanoma	13.6% (PR: 6.8–25.0)	1.34	41.2% (37.8–44.8)	5.6% (2.8–10.3)
Colorectal cancer	9.6% (PR: 8.4–47.7)	1.87	57.4% (52.7–62.4)	5.5% (4.8–27.6)
Breast cancer	8.9% (PR: 8.5–26.3)	1.61	49.4% (45.3–53.7)	4.4% (4.2–13.1)
Prostate cancer	2.4% (PR: 1.7–13.8)	1.70	52.2% (47.9–56.7)	1.2% (0.9–7.3)
Top 5 cancers subtotal	11.1% (PPR: 10.1–20.9)	1.82	56.0% (52.3–58.8)	6.2% (5.5–11.7)
Other cancers	11.1% (PPR: 10.1–20.9)[b]	2.13	65.5% (60.1–71.1)	7.3% (6.6–13.9)
Total cancers	11.1% (PPR: 10.1–20.9)	1.95	59.9% (56.7–62.7)	6.6% (6.0–12.6)
Additional totals				
Total Big Three (top 5 only)	9.7% (PPR: 8.2–12.3)	1.75	53.9% (52.7–56.8)	5.2% (4.5–6.7)
Total Big Three (top 5 + other)	9.6% (PPR: 8.0–12.2)	1.30	39.9% (39.6–43.2)	3.8% (3.3–5.1)

[a]Shown are either 95% confidence intervals (CIs), plausible ranges (PRs), or probabilistic plausible ranges (PPRs). We used PRs when there was heterogeneity in the findings across disease-specific studies of similar quality or when two different error rates were defined within a single study based on different lengths of diagnostic delay (see Tables 1–3 footnotes for additional details). PPRs derive from Monte Carlo analysis. [b]We made the simplifying assumption that error rates for "Other" (unnamed) diseases in each category would be similar to those for the top 5 in that same category. Thus, within-category miss rates for "Other" diseases represent means from the top 5 conditions. These means considered disease incidence (e.g. myocardial infarction had proportionally more impact on the final mean than aortic aneurysm and dissection, because there are many more incident cases of myocardial infarction) (see Supplementary Material A2). PPRs derive from Monte Carlo analysis. CI, confidence interval; PPR, probabilistic plausible range; PR, plausible range.

to disease-specific studies we found that assessed harm rates per error. For instance, a study of missed aneurysmal subarachnoid hemorrhages found that 50.9% (n = 27/53) had died or were permanently disabled at 1 year [70]; our estimate for missed stroke (which included missed subarachnoid hemorrhage) was 54.7%. Similarly, a study of missed spinal abscess found that 60.6% (n = 40/66) had died or suffered severe harm [23]; our estimate was 57.4%.

More importantly, we were able to approximate combined error-harm rates from the literature for three of our vascular conditions, and these were consistent with our study outputs (Table 6). Specifically, a recent analysis of Medicare claims data by Waxman et al. on emergency department discharges of major vascular events (acute myocardial infarction, aortic aneurysm rupture, aortic dissection, stroke, and subarachnoid hemorrhage) produced very similar results to our analysis [71]. This retrospective cohort study linked emergency department discharges to subsequent hospitalizations in order to identify short-term adverse events related to presumed missed diagnosis. They estimated that 3.9% (range across diseases 2.3%–4.5%) of dangerous vascular events involved an

Table 6: Comparison of present study estimates of harm rates per incident case of disease to prior literature [71].

Condition	Present study	Prior literature	Notes
Stroke	4.8%	4.1%	Both our current estimate and that of Waxman et al. [71] include subarachnoid hemorrhage within the stroke disease grouping. Although Waxman et al. did not report stroke severity, it is known that strokes were severe enough to prompt both a return to the emergency department and to warrant hospital admission (i.e. stroke-related adverse events). Risk of mortality rises nearly 5-fold after an initial miss [70].
Myocardial infarction	1.2%	2.3%	The most likely explanation for the disparity between these two estimates is that Waxman et al.'s cohort [71] included *any* hospital admission for myocardial infarction, even if the outcome was not serious morbidity or mortality. Thus, our current estimate is likely lower because it reflects only the more serious harms.
Aortic aneurysm and dissection	16.8%	4.0%	Waxman et al.'s methodology [71] relies on hospital admissions after emergency department *discharge*, so does not account for within-visit delays in diagnosis. The mortality with aortic dissection rises by ~1% per hour and median time to diagnosis is ~4 h [45]. Delays >4 h occur in ~50%, and delays >24 h in ~25% of cases [45]. Thus, the difference between the two results is almost certainly explained by the lack of accounting by Waxman et al. for increases in misdiagnosis-related harms *within* a single visit or hospitalization. This difference could easily account for a 10–15% difference in serious morbidity and mortality.

NCI SEER, National Cancer Institute Surveillance, Epidemiology, and End Results program/database.

"observed above expected" recent treat-and-release emergency department visit antecedent to a hospital admission for the vascular event in question. Although framed by the authors of that article as "diagnostic errors", their methods more closely approximate misdiagnosis-related harms [32]. Only aortic aneurysm/dissection had a very different measured per-incident disease case harm rate, and this was likely because within-visit or within-hospital delays (e.g. 12 h) go unaccounted for when using revisit-based analyses (see Table 6 notes).

Discussion

This study, based on an analysis of previously published literature, provides the first robust estimates of diagnostic error and serious misdiagnosis-related harm rates for 15 diseases that result in nearly half of all the permanent disability or death due to diagnostic error [18]. Our results were bolstered by the use of high-quality prior literature; expert feedback and validation; and, where possible, corroboration via comparison to studies using other methods. These findings are important because they can be used immediately to develop national estimates of aggregate harms from diagnostic error; furthermore, they allow us to target diagnostic improvement initiatives to diseases with the highest error and harm rates.

Disease-specific diagnostic error rates ranged from 2.2% to 62.1% and combined diagnostic error-harm rates ranged from 1.2% to 35.6%. For vascular events and infections, the lowest error and error-harm rates were seen with the most common dangerous diseases, such as myocardial infarction (2.2%, 1.2%), stroke (8.7%, 4.8%), pneumonia (9.5%, 4.5%), and sepsis (9.5%, 5.4%). Conversely, the highest error and error-harm rates were seen with the least common diseases, such as endocarditis (25.5%, 13.4%), meningitis and encephalitis (25.6%, 14.3%), aortic aneurysm and dissection (27.9%, 16.8%), and spinal abscess (62.1%, 35.6%). Presumably this disease-incidence dependence of error frequency and harms relates to a combination of factors for clinicians, such as having less overall experience in assessment of patients with low-prevalence diseases and presentations or, conversely, choosing not to pursue low-likelihood diagnoses because of known low baseline prevalence. Both would make sense, given the scant feedback to calibrate physician diagnostic skills even for some of the more common, dangerous conditions such as stroke [72] and the relatively low per-symptom disease prevalence in clinical scenarios where the cases are most often missed (e.g. only 3–5% of acute dizziness is from stroke [73]).

For cancers, the variation in error and harm frequency may have correlated more with public awareness and utilization of screening programs than purely disease incidence. The lowest rate was seen with prostate cancer (2.4%, 1.2%), where frequent screening (even overuse) has generally become the US norm [74]; the highest rate was seen with lung cancer (22.5, 13.8%), where, despite public health efforts, screening remains below the recommended levels [75]. It is also possible that at least some of the differences in harms could be attributable to treatment

differences (i.e. fewer effective therapies for later-stage lung cancers contribute to the adverse impact of diagnostic delays) (Supplementary Material B2).

Despite a small but steady decline in hospital autopsy-determined diagnostic errors over the past several decades [37], we found no evidence from the literature that overall rates were declining appreciably over time. This comports with a recent study analyzing Medicare data (2007–2014) which showed that the risks of missed major vascular events was either stable (myocardial infarction, aortic dissection) or rising (stroke, subarachnoid hemorrhage, aortic aneurysm rupture) [71]. Why misdiagnosis of some diseases might be rising over time is unclear, but this trend is alarming and deserves greater attention going forward.

A key methodological insight from this study is that per-error harm rates are tightly hinged to the threshold for considering a diagnostic delay an error (Supplementary Material B1). This was apparent both in the review of available literature and during feedback from clinical experts. Of course, it is unsurprising that trivial delays (e.g. <72 h for a cancer diagnosis, even if malignant) will be nearly universal, yet devoid of impact on patient outcomes; likewise, it is also no surprise that very long delays (e.g. >1 year for a malignant cancer diagnosis) will be uncommon, yet highly impactful. A clinically meaningful delay is a function of underlying disease biology and natural history – for colorectal cancer, delays up to ~6–9 months likely have no impact [62]; for aortic dissection, minutes probably count [45].

Importantly, however, this fact makes it challenging to combine data from disparate sources about error and harms, as definitions may not align. We were only able to do so here with confidence because of independent corroboration of these disease-specific estimates from two sources – clinical experts, demonstrating face validity, and a large independent study by Waxman et al. [71] that used very different methods, demonstrating construct validity [76]. In the latter case, our final error-harm rates for stroke and heart attack were similar to those from the Waxman et al.'s study, demonstrating convergent validity, while those for aortic aneurysm and dissection differed substantially, as would be expected based on methodological differences, demonstrating divergent validity (Table 6). Thus, future estimates of misdiagnosis-related harms would benefit from disease-specific study designs that measure error and case-mix-adjusted harms in one study [70] or address harms directly (e.g. SPADE [32]). Ideally, these would also address issues of treatment advances and case-mix adjustment (Supplementary Material B2 and B3).

Finally, it is worth pointing out that diseases historically receiving the most sustained attention to diagnosis (i.e. research funding, clinical quality improvement, public awareness) are the ones with the lowest harm rates. Myocardial infarction is the prototype and the only acute illness approaching the target "standard" of <1% harmed often cited in the emergency department [77]. This is, of course, after a half century of focused efforts to automate electrocardiogram interpretation [78], develop and refine biomarkers (e.g. troponin) [79], and create routine diagnostic protocols for chest pain or suspected acute coronary syndromes. Similarly, basic research studies and clinical trials focused on prostate cancer biomarkers (e.g. prostate-specific antigen) date back to the 1960s [80]. Achieving similar gains may be possible for other key diseases, but only if we make sustained investments in improving diagnosis (e.g. missed stroke [31] in acute dizziness [73], where novel bedside tests [81] and tele-medicine [82] have shown early promise).

Limitations

This study is limited by the quality of available literature on diagnostic errors and harms. Our estimates of diagnostic error rates for infections were based on smaller sample sizes than those in the other two "Big Three" categories, so are generally less precise. Not all studies were US based and recent, so current rates could differ – however, at least one study [45] found higher error rates in North America (mostly US) than Europe, and several studies assessed for trends toward lower error rates in recent years and found only stable or worsening diagnostic accuracy. Our main estimates of harms were derived from generic, disease-agnostic studies, then weighted based on malpractice claims severity to make them disease-specific; the weights themselves may be inexact, but our final combined error-harm estimates were face valid to experts. Although social desirability bias could have influenced feedback from domain experts, findings were concordant with previously published literature, where available (Table 6).

Harm rates reflect only delayed or missed diagnoses (i.e. false-negative dangerous disease diagnoses), so do not account for harms from treatment for wrong diagnoses (i.e. false-positive dangerous disease diagnoses); thus, for example, any harms associated with thrombolytic therapy for presumed ischemic stroke in a patient who actually has migraine with aura (i.e. does *not* actually have stroke) are not considered here. We also did not consider the cumulative morbidity of less serious (but more frequent) harms from diagnostic error (e.g. pain, temporary disability, psychological distress). Most of the studies cited did not consider communication failures with patients, so these NAM-defined diagnostic errors were not fully accounted for in the current estimates. Thus, total misdiagnosis-related harms are likely greater than assessed here. It is unknown if the harms represented in this analysis would necessarily

have been prevented by prompt, correct diagnosis, and, in some cases, attempting to reduce false negatives could have adverse, unintended consequences from false positives.

Conclusions

We estimate that roughly one in 10 patients with a dangerous "Big Three" disease is misdiagnosed, and roughly half of those misdiagnosed die or are permanently disabled as a result. Diagnostic error and harm rates vary substantially across dangerous diseases and do not appear to be declining over time. For a given disease, error and harm rates are inversely related and probably tightly coupled – this makes estimates of combined error-harm rates per incident disease case more stable, more comparable across studies, and more clinically relevant than either quantity alone. The lowest error and harm rates were seen with extensively researched conditions that have received sustained attention and investments to improve diagnosis over several decades. These findings will immediately facilitate creation of national estimates of aggregate harms from diagnostic error. Simultaneously, they should also help guide and focus future diagnostic improvement initiatives toward conditions where current diagnostic performance is lacking.

Acknowledgments: We would like to thank Drs. John Ely, Gordon Schiff, and Laura Zwaan for providing unpublished details from their diagnostic errors research studies that supported our analyses. In addition, we are grateful for input from the experts listed below, who provided input on the face validity of diagnostic error and harm rates for the 15 specific "Big Three" diseases that fell within their areas of clinical domain expertise, and expressed their willingness to be acknowledged in the two associated manuscripts. Their names are listed alphabetically within each "Big Three" category to which their expertise was applied:

Vascular events:
- Michael Bowdish (thoracic surgery, focus on aortic dissection)
- Robbin Cohen (thoracic surgery, focus on aortic dissection)
- Jonathan Edlow (emergency medicine, focus on diagnostic errors and neurologic emergencies)
- Joshua Goldstein (emergency medicine, focus on neurologic emergencies, especially stroke)
- Elliott Haut (surgery, focus on venous thromboembolism, quality and safety)
- William Meurer (emergency medicine, focus on neurologic emergencies, health services research)
- Julie Miller (cardiology, focus on myocardial infarction, quality and safety)
- Rodney Omron (emergency medicine, focus on diagnostic errors)
- Susan Peterson (emergency medicine, focus on diagnostic errors, quality and safety)

Infections:
- Paul Auwaerter (infectious disease, focus on atypical and chronic infections)
- Justin McArthur (neurology, focus on neurologic infections)
- Richard Rothman (emergency medicine, focus on acute infections)
- Jenny Townsend (infectious disease, focus on antibiotic overuse, quality and safety)
- Arun Venkatesan (neurology, focus on neurologic infections)
- Jonathan Zenilman (infectious disease, focus on hospital infections, quality and safety)

Cancers:
- Michael Carducci (oncology, focus on prostate cancer)
- Ross Donehower (oncology, focus on colorectal cancer)
- Josephine Feliciano (oncology, focus on lung cancer)
- Russell Hales (radiation oncology, focus on lung cancer, quality and safety)
- Daniel Laheru (oncology, focus on colorectal, pancreatic cancer)
- Art Papier (dermatology, focus on melanoma, diagnostic errors)
- Antonio Wolff (oncology, focus on breast cancer)

Author contributions: David Newman-Toker: I declare that I designed the study; had primary oversight over the data analysis; conducted diagnostic error 95% CI calculations; designed the figures; authored the primary manuscript draft and all major revisions; and that I have seen and approved the final version. I serve as an unpaid member of the Board of Directors of the Society to Improve Diagnosis in Medicine, and as its President. I periodically serve as a medico-legal consultant for both plaintiff and defense in cases related to diagnostic error. I have no other relevant conflicts of interest. **Zheyu Wang**: I declare that I led all statistical analyses; edited the manuscript for scientific content; and that I have seen and approved the final version. I have no conflicts of interest. **Yuxin Zhu**: I declare that I assisted in design and conduct of statistical analyses; edited the manuscript for scientific content;

and that I have seen and approved the final version. I have no conflicts of interest. **Najlla Nassery**: I declare that I assisted in study design; edited the manuscript for scientific content; and that I have seen and approved the final version. I have no conflicts of interest. **Ali Saber Tehrani**: I declare that I assisted in study conduct; edited the manuscript for scientific content; and that I have seen and approved the final version. I have no conflicts of interest. **Adam Schaffer**: I declare that I assisted in study design; assisted in the analysis of malpractice data, including case reviews; edited the manuscript for scientific content; and that I have seen and approved the final version. I have no conflicts of interest. **Chihwen Winnie Yu-Moe**: I declare that I conducted the data analysis of malpractice data; edited the manuscript for scientific content; and that I have seen and approved the final version. I have no conflicts of interest. **Gwendolyn Clemens**: I declare that I assisted in study conduct; edited the manuscript for scientific content; and that I have seen and approved the final version. I have no conflicts of interest. **Mehdi Fanai**: I declare that I assisted in data analysis; edited the manuscript for scientific content; and that I have seen and approved the final version. I have no conflicts of interest. **Dana Siegal**: I declare that I assisted in study design; oversaw the analysis of CRICO malpractice data; edited the manuscript for scientific content; and that I have seen and approved the final version. I serve as an unpaid member of the Board of Directors of the Society to Improve Diagnosis in Medicine. The first/corresponding author (David E. Newman-Toker) had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The first/corresponding author also had final responsibility for the decision to submit for publication. All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: The Society to Improve Diagnosis in Medicine, through a grant from the Gordon & Betty Moore Foundation. Dr. Newman-Toker's effort was supported partly by the Armstrong Institute Center for Diagnostic Excellence at the Johns Hopkins University School of Medicine.

Employment or leadership: Dr. Newman-Toker conducts research related to diagnostic error, including serving as the principal investigator for grants on this topic. He serves as an unpaid member of the Board of Directors of the Society to Improve Diagnosis in Medicine and as its current President. He serves as a medico-legal consultant for both plaintiff and defense in cases related to diagnostic error. Dana Siegal serves as an unpaid member of the Board of Directors of the Society to Improve Diagnosis in Medicine. There are no other conflicts of interest. None of the authors have any financial or personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Role of medical writer or editor: No medical writer or editor was involved in the creation of this manuscript.

Previous presentation of the information reported in the manuscript: None. Contents were accepted for publication in abstract form for the Diagnostic Error in Medicine meeting in New Orleans, LA, November 4–6, 2018. The main presentation of results was withdrawn in anticipation of a future press event around the final, published study results, but aspects of the study results were presented in poster form.

Persons who have made substantial contributions to the work but are not authors: None.

References

1. Improving Diagnosis in Healthcare. Institute of Medicine, 2015. http://www.nationalacademies.org/hmd/Reports/2015/Improving-Diagnosis-in-Healthcare.aspx. Accessed 9 Feb 2020.
2. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013;22 Suppl 2:ii21–7.
3. Zwaan L, de Bruijne M, Wagner C, Thijs A, Smits M, van der Wal G, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. Arch Intern Med 2010;170:1015–21.
4. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. JAMA Intern Med 2013;173:418–25.
5. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23:727–31.
6. Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. Am J Med 2000;108:642–9.
7. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. J Gen Intern Med 2003;18:61–7.
8. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. J Am Med Assoc 2001;286:415–20.
9. Weingart SN, Davis RB, Palmer RH, Cahalane M, Hamel MB, Mukamal K, et al. Discrepancies between explicit and implicit review: physician and nurse assessments of complications and quality. Health Serv Res 2002;37:483–98.
10. Wears RL, Nemeth CP. Replacing hindsight with insight: toward better understanding of diagnostic failures. Ann Emerg Med 2007;49:206–9.

11. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. Med Care 2004;42:1066–72.
12. Ambulatory Care Use and Physician Office Visits (2016). Centers for Disease Control and Prevention, National Center for Health Statistics. January 2017. http://www.cdc.gov/nchs/fastats/physician-visits.htm. Accessed 9 Feb 2020.
13. Emergency Department Visits (2016). Centers for Disease Control and Prevention, National Center for Health Statistics. January 2017. http://www.cdc.gov/nchs/fastats/emergency-department.htm. Accessed 9 Feb 2020.
14. HCUP Fast Stats – Trends in Inpatient Stays (2007–2016). Healthcare Cost and Utilization Project (HCUP). December 2019. Agency for Healthcare Research & Quality, Rockville, MD. https://www.hcup-us.ahrq.gov/faststats/NationalTrendsServlet. Accessed 9 Feb 2020.
15. Sonderegger-Iseli K, Burger S, Muntwyler J, Salomon F. Diagnostic errors in three medical eras: a necropsy study. Lancet 2000;355:2027–31.
16. Newman-Toker DE, Tucker L, on behalf of the Society to Improve Diagnosis in Medicine Policy Committee. Roadmap for Research to Improve Diagnosis, Part 1: Converting National Academy of Medicine Recommendations into Policy Action: Society to Improve Diagnosis in Medicine; 2018. https://www.improvediagnosis.org/roadmap/. Accessed 9 Feb 2020.
17. Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, 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;22:672–80.
18. Newman-Toker DE, Schaffer AC, Yu-Moe CW, Nassery N, Saber Tehrani AS, Clemens GD, et al. Serious misdiagnosis-related harms in malpractice claims: the "Big Three" – vascular events, infections, and cancers. Diagnosis (Berl) 2019;6:227–40.
19. Deaths and Mortality (2017). Centers for Disease Control and Prevention, National Center for Health Statistics. May 2017. http://www.cdc.gov/nchs/fastats/deaths.htm. Accessed 9 Feb 2020.
20. Berlin L. Accuracy of diagnostic procedures: has it improved over the past five decades? AJR Am J Roentgenol 2007;188:1173–8.
21. Calder L, Pozgay A, Riff S, Rothwell D, Youngson E, Mojaverian N, et al. Adverse events in patients with return emergency department visits. BMJ Qual Saf 2015;24:142–8.
22. Davis DP, Wold RM, Patel RJ, Tran AJ, Tokhi RN, Chan TC, et al. The clinical presentation and impact of diagnostic delays on emergency department patients with spinal epidural abscess. J Emerg Med 2004;26:285–91.
23. Bhise V, Meyer AN, Singh H, Wei L, Russo E, Al-Mutairi A, et al. Errors in diagnosis of spinal epidural abscesses in the era of electronic health records. Am J Med 2017;130:975–81.
24. Kerber KA, Morgenstern LB, Meurer WJ, McLaughlin T, Hall PA, Forman J, et al. Nystagmus assessments documented by emergency physicians in acute dizziness presentations: a target for decision support? Acad Emerg Med 2011;18:619–26.
25. Herzog R, Elgort DR, Flanders AE, Moley PJ. Variability in diagnostic error rates of 10 MRI centers performing lumbar spine MRI examinations on the same patient within a 3-week period. Spine J 2017;17:554–61.
26. Elliott CG, Goldhaber SZ, Jensen RL. Delays in diagnosis of deep vein thrombosis and pulmonary embolism. Chest 2005;128:3372–6.
27. Troxel DB. Diagnostic Error in Medical Practice by Specialty. The Doctor's Advocate 2014;2:5.
28. Hanscom R, Small M, Lambrecht A. Diagnostic accuracy: room for improvement: coverys; 2018. https://coverys.com/PDFs/Coverys_Diagnostic_Accuracy_Report.aspx. Accessed 9 Feb 2020.
29. Winters B, Custer J, Galvagno Jr SM, Colantuoni E, Kapoor SG, Lee H, et al. Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. BMJ Qual Saf 2012;21:894–902.
30. Custer JW, Winters BD, Goode V, Robinson KA, Yang T, Pronovost PJ, et al. Diagnostic errors in the pediatric and neonatal ICU: a systematic review. Pediatr Crit Care Med 2015;16:29–36.
31. Tarnutzer AA, Lee SH, Robinson KA, Wang Z, Edlow JA, Newman-Toker DE. ED misdiagnosis of cerebrovascular events in the era of modern neuroimaging: a meta-analysis. Neurology 2017;88:1468–77.
32. 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.
33. Newman-Toker DE, Pronovost PJ. Diagnostic errors – the next frontier for patient safety. J Am Med Assoc 2009;301:1060–2.
34. Newman-Toker DE. A unified conceptual model for diagnostic errors: underdiagnosis, overdiagnosis, and misdiagnosis. Diagnosis (Berl) 2014;1:43–8.
35. NAIC Malpractice Claims, Final Compilation. Brookfield, WI: National Association of Insurance Commissioners; 1980. https://www.naic.org/documents/prod_serv_special_med_lb.pdf. Accessed 9 Feb 2020.
36. Guideline for Implementation of Medical Professional Liability Closed Claim Reporting (GDL-1077). National Association of Insurance Commissioners; 2010. http://www.naic.org/store/free/GDL-1077.pdf. Accessed 9 Feb 2020.
37. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. J Am Med Assoc 2003;289:2849–56.
38. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med 2009;169:1881–7.
39. Ely JW, Kaldjian LC, D'Alessandro DM. Diagnostic errors in primary care: lessons learned. J Am Board Fam Med 2012;25:87–97.
40. Okafor N, Payne VL, Chathampally Y, Miller S, Doshi P, Singh H. Using voluntary reports from physicians to learn from diagnostic errors in emergency medicine. Emerg Med J 2016;33:245–52.
41. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970;57:97–109.
42. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573–7.
43. Pope JH, Aufderheide TP, Ruthazer R, Woolard RH, Feldman JA, Beshansky JR, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. N Engl J Med 2000;342:1163–70.
44. Azhar B, Patel SR, Holt PJ, Hinchliffe RJ, Thompson MM, Karthikesalingam A. Misdiagnosis of ruptured abdominal aortic aneurysm: systematic review and meta-analysis. J Endovasc Ther 2014;21:568–75.
45. Harris KM, Strauss CE, Eagle KA, Hirsch AT, Isselbacher EM, Tsai TT, et al. Correlates of delayed recognition and treatment

of acute type A aortic dissection: the International Registry of Acute Aortic Dissection (IRAD). Circulation 2011;124:1911–8.
46. Kassahun WT, Schulz T, Richter O, Hauss J. Unchanged high mortality rates from acute occlusive intestinal ischemia: six year review. Langenbecks Arch Surg 2008;393:163–71.
47. Eltarawy IG, Etman YM, Zenati M, Simmons RL, Rosengart MR. Acute mesenteric ischemia: the importance of early surgical consultation. Am Surg 2009;75:212–9.
48. Firetto MC, Lemos AA, Marini A, Avesani EC, Biondetti PR. Acute bowel ischemia: analysis of diagnostic error by overlooked findings at MDCT angiography. Emerg Radiol 2013;20:139–47.
49. Lehtimaki TT, Karkkainen JM, Saari P, Manninen H, Paajanen H, Vanninen R. Detecting acute mesenteric ischemia in CT of the acute abdomen is dependent on clinical suspicion: review of 95 consecutive patients. Eur J Radiol 2015;84: 2444–53.
50. Vaillancourt S, Guttmann A, Li Q, Chan IY, Vermeulen MJ, Schull MJ. Repeated emergency department visits among children admitted with meningitis or septicemia: a population-based study. Ann Emerg Med 2015;65:625–32 e3.
51. Scott HF, Greenwald EE, Bajaj L, Deakyne Davies SJ, Brou L, Kempe A. The sensitivity of clinician diagnosis of sepsis in tertiary and community-based emergency settings. J Pediatr 2018;195:220–7 e1.
52. Morr M, Lukasz A, Rubig E, Pavenstadt H, Kumpers P. Sepsis recognition in the emergency department – impact on quality of care and outcome? BMC Emerg Med 2017;17:11.
53. Rhee C, Jones TM, Hamad Y, Pande A, Varon J, O'Brien C, et al. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. JAMA Netw Open 2019;2:e187571.
54. McIntyre PB, Macintyre CR, Gilmour R, Wang H. A population based study of the impact of corticosteroid therapy and delayed diagnosis on the outcome of childhood pneumococcal meningitis. Arch Dis Child 2005;90:391–6.
55. Claessens YE, Debray MP, Tubach F, Brun AL, Rammaert B, Hausfater P, et al. Early chest computed tomography scan to assist diagnosis and guide treatment decision for suspected community-acquired pneumonia. Am J Respir Crit Care Med 2015;192:974–82.
56. Zwaan L, Thijs A, Wagner C, Timmermans DR. Does inappropriate selectivity in information use relate to diagnostic errors and patient harm? The diagnosis of patients with dyspnea. Soc Sci Med 2013;91:32–8.
57. N'Guyen Y, Duval X, Revest M, Saada M, Erpelding ML, Selton-Suty C, et al. Time interval between infective endocarditis first symptoms and diagnosis: relationship to infective endocarditis characteristics, microorganisms and prognosis. Ann Med 2017;49:117–25.
58. Nadpara P, Madhavan SS, Tworek C. Guideline-concordant timely lung cancer care and prognosis among elderly patients in the United States: a population-based study. Cancer Epidemiol 2015;39:1136–44.
59. Vidaver RM, Shershneva MB, Hetzel SJ, Holden TR, Campbell TC. Typical time to treatment of patients with lung cancer in a multisite, US-based study. J Oncol Pract 2016;12:e643–53.
60. Singh H, Hirani K, Kadiyala H, Rudomiotov O, Davis T, Khan MM, et al. Characteristics and predictors of missed opportunities in lung cancer diagnosis: an electronic health record-based study. J Clin Oncol 2010;28:3307–15.
61. Partridge AH, Hughes ME, Ottesen RA, Wong YN, Edge SB, Theriault RL, et al. The effect of age on delay in diagnosis and stage of breast cancer. Oncologist 2012;17:775–82.
62. Corley DA, Jensen CD, Quinn VP, Doubeni CA, Zauber AG, Lee JK, et al. Association between time to colonoscopy after a positive fecal test result and risk of colorectal cancer and cancer stage at diagnosis. J Am Med Assoc 2017;317: 1631–41.
63. Pruitt SL, Harzke AJ, Davidson NO, Schootman M. Do diagnostic and treatment delays for colorectal cancer increase risk of death? Cancer Cause Control 2013;24:961–77.
64. Redaniel MT, Martin RM, Ridd MJ, Wade J, Jeffreys M. Diagnostic intervals and its association with breast, prostate, lung and colorectal cancer survival in England: historical cohort study using the Clinical Practice Research Datalink. PLoS One 2015;10:e0126608.
65. Baade PD, Youl PH, English DR, Mark Elwood J, Aitken JF. Clinical pathways to diagnose melanoma: a population-based study. Melanoma Res 2007;17:243–9.
66. Baade PD, English DR, Youl PH, McPherson M, Elwood JM, Aitken JF. The relationship between melanoma thickness and time to diagnosis in a large population-based study. Arch Dermatol 2006;142:1422–7.
67. Strazzulla LC, Li X, Zhu K, Okhovat JP, Lee SJ, Kim CC. Clinicopathologic, misdiagnosis, and survival differences between clinically amelanotic melanomas and pigmented melanomas. J Am Acad Dermatol 2019;80:1292–8.
68. Thomas NE, Kricker A, Waxweiler WT, Dillon PM, Busman KJ, From L, et al. Comparison of clinicopathologic features and survival of histopathologically amelanotic and pigmented melanomas: a population-based study. JAMA Dermatol 2014;150:1306–314.
69. Moy E, Barrett M, Coffey R, Hines AL, Newman-Toker DE. Missed diagnoses of acute myocardial infarction in the emergency department: variation by patient and facility characteristics. Diagnosis (Berl) 2015;2:29–40.
70. Kowalski RG, Claassen J, Kreiter KT, Bates JE, Ostapkovich ND, Connolly ES, et al. Initial misdiagnosis and outcome after subarachnoid hemorrhage. J Am Med Assoc 2004;291:866–9.
71. Waxman DA, Kanzaria HK, Schriger DL. Unrecognized cardiovascular emergencies among medicare patients. JAMA Intern Med 2018;178:477–84.
72. Omron R, Kotwal S, Garibaldi BT, Newman-Toker DE. The diagnostic performance feedback "calibration gap": why clinical experience alone is not enough to prevent serious diagnostic errors. AEM Educ Train 2018;2:339–42.
73. Newman-Toker DE. Missed stroke in acute vertigo and dizziness: it is time for action, not debate. Ann Neurol 2016;79:27–31.
74. Khairnar R, Mishra MV, Onukwugha E. Impact of United States Preventive Services Task Force recommendations on utilization of prostate-specific antigen screening in medicare beneficiaries. Am J Clin Oncol 2018. doi:10.1097/COC.0000000000000431. [Epub ahead of print].
75. Richards TB, Doria-Rose VP, Soman A, Klabunde CN, Caraballo RS, Gray SC, et al. Lung cancer screening inconsistent with U.S. Preventive Services Task Force recommendations. Am J Prev Med 2019;56:66–73.
76. Karras DJ. Statistical methodology: II. Reliability and validity assessment in study design, Part B. Acad Emerg Med 1997;4:144–7.

77. Newman-Toker DE, Edlow JA. High-stakes diagnostic decision rules for serious disorders: the Ottawa subarachnoid hemorrhage rule. J Am Med Assoc 2013;310:1237–9.
78. Macfarlane PW. A brief history of computer-assisted electrocardiography. Methods Inf Med 1990;29:272–81.
79. Garg P, Morris P, Fazlanie AL, Vijayan S, Dancso B, Dastidar AG, et al. Cardiac biomarkers of acute coronary syndrome: from history to high-sensitivity cardiac troponin. Intern Emerg Med 2017;12:147–55.
80. Catalona WJ. History of the discovery and clinical translation of prostate-specific antigen. Asian J Urol 2014;1:12–4.
81. Newman-Toker DE, Curthoys IS, Halmagyi GM. Diagnosing stroke in acute vertigo: the HINTS family of eye movement tests and the future of the "Eye ECG". Semin Neurol 2015;35:506–21.
82. Gold D, Peterson S, McClenney A, Tourkevich R, Brune A, Choi W, et al. Diagnostic impact of a device-enabled remote "Tele-Dizzy" consultation service [abstract]. Diagnostic Error in Medicine, 12th Annual Conference (Washington, DC). November 10–13, 2019.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/dx-2019-0104).