Burden of serious harms from diagnostic error in the USA

David E Newman-Toker ,[1,2] Najlla Nassery,[3] Adam C Schaffer,[4,5] Chihwen Winnie Yu-Moe,[5] Gwendolyn D Clemens,[6] Zheyu Wang,[6,7] Yuxin Zhu,[1,6] Ali S. Saber Tehrani,[1] Mehdi Fanai,[1] Ahmed Hassoon,[1,2] Dana Siegal[8,9]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjqs-2021-014130).

For numbered affiliations see end of article.

Correspondence to
Dr David E Newman-Toker, Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA; toker@jhu.edu

Received 18 August 2021
Accepted 24 June 2023
Published Online First 17 July 2023

► http://dx.doi.org/10.1136/bmjqs-2023-016496

To cite: Newman-Toker DE, Nassery N, Schaffer AC, *et al*. *BMJ Qual Saf* 2024;**33**:109–120.

ABSTRACT

Background Diagnostic errors cause substantial preventable harms worldwide, but rigorous estimates for total burden are lacking. We previously estimated diagnostic error and serious harm rates for key dangerous diseases in major disease categories and validated plausible ranges using clinical experts.

Objective We sought to estimate the annual US burden of serious misdiagnosis-related harms (permanent morbidity, mortality) by combining prior results with rigorous estimates of disease incidence.

Methods Cross-sectional analysis of US-based nationally representative observational data. We estimated annual incident vascular events and infections from 21.5 million (M) sampled US hospital discharges (2012–2014). Annual new cancers were taken from US-based registries (2014). Years were selected for coding consistency with prior literature. Disease-specific incidences for 15 major vascular events, infections and cancers ('Big Three' categories) were multiplied by literature-based rates to derive diagnostic errors and serious harms. We calculated uncertainty estimates using Monte Carlo simulations. Validity checks included sensitivity analyses and comparison with prior published estimates.

Results Annual US incidence was 6.0 M vascular events, 6.2 M infections and 1.5 M cancers. Per 'Big Three' dangerous disease case, weighted mean error and serious harm rates were 11.1% and 4.4%, respectively. Extrapolating to all diseases (including non-'Big Three' dangerous disease categories), we estimated total serious harms annually in the USA to be 795 000 (plausible range 598 000–1 023 000). Sensitivity analyses using more conservative assumptions estimated 549 000 serious harms. Results were compatible with setting-specific serious harm estimates from inpatient, emergency department and ambulatory care. The 15 dangerous diseases accounted for 50.7% of total serious harms and the top 5 (stroke, sepsis, pneumonia, venous thromboembolism and lung cancer) accounted for 38.7%.

Conclusion An estimated 795 000 Americans become permanently disabled or die annually across care settings because dangerous diseases are misdiagnosed. Just 15 diseases account for about half of all serious harms, so the problem may be more tractable than previously imagined.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Diagnostic errors are known to be common, costly and often catastrophic in their health outcomes for patients.
⇒ Nevertheless, current estimates of the aggregate burden of serious harms resulting from medical misdiagnosis vary widely.

WHAT THIS STUDY ADDS

⇒ This study provides the first national estimate of permanent morbidity and mortality resulting from diagnostic errors across all clinical settings, including both hospital-based and clinic-based care (0.6–1.0 million each year in the USA alone).
⇒ It does so via an approach that extrapolates from disease-based estimates for the most common dangerous conditions that often cause serious harms when missed—vascular events, infections and cancers.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Because the overall burden of serious misdiagnosis-related harms is quite large, improving diagnosis of dangerous diseases most often responsible—stroke, sepsis, pneumonia, venous thromboembolism and lung cancer—constitutes an urgent public health imperative.

INTRODUCTION

Diagnostic error is a major source of preventable harms worldwide across clinical settings,[1–6] but epidemiologically valid estimates of overall misdiagnosis-related morbidity and mortality are lacking. The US National Academy of Medicine describes improving diagnosis in healthcare as a 'moral, professional, and public health imperative'.[7] In its 2015 report, the National Academy concluded that 'most people will experience at least

one diagnostic error in their lifetime, sometimes with devastating consequences'. However, the report also noted 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'.[7] This concern is reflected in the wide variation of US estimates for total annual diagnostic errors (12 million (M) to >100 M) and serious misdiagnosis-related harms (40 000 to 4 M).[8] No studies have yet used nationally representative datasets to measure aggregate US diagnostic errors or harms.

Given wide variation in prior estimates of total diagnostic errors and harms,[8 9] we pursued a novel disease-based approach to constructing a national estimate that would span ambulatory clinic, emergency department and inpatient care. The disease-based approach leveraged three major disease categories—vascular events, infections and cancers (the 'Big Three')—found in both malpractice claims and clinical studies of diagnostic error to account for three-quarters of serious harms.[9] To estimate the total US burden of medical misdiagnosis, we multiplied national estimates of disease incidence (including those initially misdiagnosed) by the disease-specific proportion of patients with that disease experiencing errors and harms. We did this for 15 key diseases causing the most harms, then extrapolated to the grand total across all diseases. To assess the robustness of our final estimates, we used sensitivity analyses to measure the impact of methodological choices and tested validity via comparison to prior literature and expert review.

METHODS

This was a three-part research study in which the first two published components[8 9] form the basis of the current analysis, which represents the third and final component (online supplemental file 1-A1). The main goal of this three-phase research project was to estimate the total number of serious misdiagnosis-related harms (ie, permanent disability or death) occurring annually in the USA across all care settings (ambulatory clinic, emergency department and inpatient). As reported previously,[8 9] each study 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 10 years of data from a large, nationwide malpractice database representing ~30% of all US claims, then comparing the proportion of 'Big Three' diseases with that from clinical practice-based (non-claims) studies[9]; (2) how common are diagnostic errors potentially causing harm among these dangerous diseases? (using estimates of error and harm rates from high-quality clinical studies,[3 8] further validated by experts) and, for this final component, (3) what is the overall epidemiological incidence of diagnostic errors and harms among these dangerous diseases? (using nationally representative databases to measure dangerous disease incidence and multiply these by error and harm rates). This final analysis also extrapolates to all (including non-'Big Three') diagnostic errors and serious misdiagnosis-related harms by using the previously reported[9] attributable fraction of 'Big Three' diseases in clinical practice. We constructed our scientific approach such that the final grand total estimates for errors and harms in the USA are based on clinical literature and US population incidence, not malpractice claims. This is because (a) no error or harm rates were taken from claims-based studies, (b) the extrapolation from 'Big Three' disease estimates to the grand total were based on the proportion of 'Big Three' diseases causing errors and harms from clinical studies (described in '*Outcome measures*' section) and (c) any impact of having used malpractice claims to construct the original disease list or weights are mathematically unrelated to the grand totals (online supplemental file 1-A2). We summarise key aspects of prior study methods[8 9] as needed for readers to follow this final component.

Diagnostic error, misdiagnosis-related harm and harm severity definitions

As reported previously,[8 9] we used published definitions for diagnostic error[7] and misdiagnosis-related harms.[10] In this study, we considered only false negative diagnoses (ie, initially missed or delayed) and associated harms.[3 8] Harms from *inappropriate use* or *overuse* of diagnostic tests,[11 12] or from *overdiagnosis* (ie, overtreatment of correctly diagnosed conditions that, left undiagnosed, would be unlikely to impact patient health)[10 13] were not considered. Harm severity was categorised according to a recognised insurance industry standard for measuring severity of injury in malpractice claims.[14 15] *Serious (high-severity) misdiagnosis-related harms* were defined as scale scores 6–9 representing serious permanent morbidity or mortality (box 1).[9]

Although technically proportions, we use the more common terminology 'rates' to describe diagnostic errors and misdiagnosis-related harms for ease of readability. The *diagnostic error rate* is the proportion of patients with a target disease who were not diagnosed in accurate and timely fashion; the *misdiagnosis-related harm rate* is the proportion of patients with a target disease who were not diagnosed in accurate and timely fashion and suffered serious harms from the target disease.

Current study design and data sources

This cross-sectional study multiplied literature-based estimates of diagnostic errors and harms (reported previously by our team[3 8]) by nationally representative epidemiological data on disease incidence (reported here for the first time) to estimate total misdiagnosis-related harms. Multiplying disease incidence by the disease-specific proportion of patients experiencing

Box 1 NAIC scale with specific exemplars used as anchors by CRICO in coding malpractice claim severity

NAIC 6—permanent significant (eg, deafness, loss of single limb, loss of eye, loss of one kidney or lung; cancers where there is a large tumour possibly with lymph node involvement—this includes cancers that are stage III and stage IV such as breast cancer with total mastectomy, lung cancer with pneumonectomy or a small cell lung cancer that is inoperable because it has already spread too far).

NAIC 7—permanent major (eg, paraplegia, blindness, loss of two limbs, brain damage).

NAIC 8—permanent grave (eg, quadriplegia, severe brain damage, lifelong care or fatal prognosis; cancer cases with distant metastasis and/or a prognosis of <6 months).

NAIC 9—death (including fetal and neonatal death).

CRICO, Controlled Risk Insurance Company; NAIC, National Association of Insurance Commissioners.

errors and harms will result in total estimates across care settings (ambulatory clinics, emergency department and inpatient). False negative diagnostic error and harm rates for 15 key diseases ((1) stroke, (2) venous thromboembolism, (3) arterial thromboembolism, (4) aortic aneurysm/dissection, (5) myocardial infarction, (6) sepsis, (7) pneumonia, (8) meningitis/encephalitis, (9) spinal abscess, (10) endocarditis, (11) lung cancer, (12) breast cancer, (13) colorectal cancer, (14) melanoma, (15) prostate cancer) were summarised from clinical studies and vetted by experts.[8] Our team published a follow-on systematic review[3] updating error rates for vascular events and infections. For the present study, we used updated rates only for diseases for which we found high-quality studies that could be subjected to formal meta-analysis[3] (diseases #1, 2, 4, 5, 6). For updated rates, we reapproached relevant experts if revised rates had >1% absolute difference and the previous point estimate fell outside the new estimate's CI. Only stroke met these criteria; we reapproached two emergency physicians and two stroke neurologists to assess the face validity of the revised rates. As reported previously, for unnamed 'other' diseases within each 'Big Three' category (ie, where it was not possible to find literature-derived rates), we substituted the average rate for that category.[8] To ensure that estimates in this final national analysis were optimised and comparable, we repeated the same statistical procedures as before[8] but using the revised error rates.

As reported previously,[3 8] diagnostic error rates were all based on studies of missed or delayed diagnoses (ie, false negatives) among patients with true disease and were abstracted from the highest quality clinical studies we could find. All studies used for these calculations had to have clinical source populations, so no malpractice or autopsy studies were included. In some cases, studies were from countries outside the USA (Australia, Canada, New Zealand, the UK and several European nations).[3 8] We discarded lower-quality studies when more rigorous studies (eg, systematic reviews, population-based sampling, large sample sizes, rigorous case ascertainment) were available. Error rates for vascular events and infections were predominantly derived from studies in emergency department or inpatient settings, while error rates for cancers were predominantly registry based.[3 8] Disease-specific misdiagnosis-related harm rates were derived by multiplying high-quality data on disease-agnostic (non-disease-specific) harms per diagnostic error (from well-respected clinical studies) by disease-specific harm-severity weights (from malpractice claims)[8] (online supplemental file 1-A2).

We derived population-based data on disease incidence from public use datasets employing nationally representative sampling or census methods. This represents the number at risk for diagnostic error across all clinical settings. All age groups were included. The annual incidence of specific conditions within the 'Big Three' disease categories (ie, vascular events, infections and cancers) was measured using discharge data from two sources: (1) the National Inpatient Sample (NIS) (2012–2014), Healthcare Cost and Utilisation Project (HCUP), Agency for Healthcare Research and Quality[16] and (2) North American Association of Central Cancer Registries (NAACCR)[17] curated by the American Cancer Society (ACS) (2014).[18] The year 2014 was chosen as the last full year in which national data were coded using the International Classification of Diseases 9th revision, Clinical Modification (ICD-9-CM), prior to the 2015 transition to the International Classification of Diseases 10th revision, Clinical Modification, for coding consistency with the previously published components of the study.[8 9]

Disease incidence data for vascular events and infections

The conservative assumption was made that incident cases of dangerous (life or limb-threatening) vascular events and infections in the USA would eventually involve a hospitalisation, even if the patient was initially misdiagnosed in an ambulatory care setting. Outpatient (eg, primary care, emergency department) visit diagnoses were *not* included separately in the disease incidence calculations because they would risk inflating disease incidence estimates through double counting. For example, if 'myocardial infarction' cases that were correctly diagnosed in outpatient care (and then later confirmed as an inpatient) had been included in the analysis, the same incident cases would be counted twice. Out-of-hospital deaths from these conditions were not considered, as cause-of-death

listings on death certificates are known to be inaccurate for some conditions (eg, myocardial infarction).[19]

HCUP NIS data were used to measure US inpatient hospital stays, counting discharge or in-hospital death diagnoses coded in either the principal or first-listed secondary diagnosis positions, as these diagnoses are often of equal, competing weight.[20] We chose this approach for the primary analysis because (1) using second-position codes can increase sensitivity without sacrificing specificity[21] and (2) 'secondary' diseases are also incident disease cases with the potential to be misdiagnosed, independent from the 'primary' disease (eg, a comorbid stroke in a patient with endocarditis might also be missed and this additional missed opportunity could also harm the patient).

Disease-level and 'Big Three' category-level code groupings were the same as those used in prior project phases[8 9] and double-checked for coherence with NIS analysis (online supplemental file 1-A3). These were derived from HCUP's Clinical Classification Software, which groups ICD-9-CM codes into clinically meaningful categories. We used NIS data (2012–2014) to estimate the annual number of hospital discharges nationwide by disease and category. A 3-year average was chosen to improve stability of incidence measures for rare conditions (eg, spinal abscess). We followed standard procedures for NIS data to derive nationally representative estimates (online supplemental file 1-A4).[22]

Disease incidence data for cancers

Inpatient hospital stays would not be a good proxy for incident cancer cases, since cancers are treated in outpatient settings and patients are usually only hospitalised for complications. Instead, national incidence counts by cancer site (ie, body location) were obtained from the 2014 ACS report.[18] As stated in the report, counts were based primarily on incidence data collected by the NAACCR, which represents 89% of the US population. ACS also used other unidentified sources to generate their final counts, but, because both NAACCR and ACS treat these registry-based estimates as a census (ie, no sampling-related uncertainty), we did the same. Some ACS categories were grouped to match the prior disease classification from earlier study phases (eg, colon and rectum cancer grouped as 'colorectal').[8 9]

Outcome measures

The main outcome measures were estimates of total annual diagnostic errors (false negatives) and serious misdiagnosis-related harms (permanent morbidity or mortality) in the USA for 2014, across all clinical settings. Outcomes were calculated for the 'Big Three' disease categories, including 15 specific diseases (ie, the previously identified[9] top five vascular events, infections and cancers), 'other' (non-top five) diseases within each category and corresponding category totals.

In turn, these 'Big Three' results were used to calculate a grand total (including non-'Big Three' dangerous diseases) using the clinical proportion of diagnostic errors (58.5%) and serious harms (75.8%) attributable to 'Big Three' diseases.[9] These proportions derive exclusively from research studies based in clinical practice (ie, not malpractice claims studies) (*see prior citation,[9] p. 237*). Mathematically, the grand total of diagnostic errors was calculated by dividing the 'Big Three' total number of diagnostic errors by 0.585. Similarly, the grand total of serious misdiagnosis-related harms was calculated by dividing the 'Big Three' total number of serious misdiagnosis-related harms by 0.758.

Using the proportion of deaths among serious harms across clinical settings (~46.7%),[6 23] we estimated total deaths (total serious harms × proportion of deaths among serious harms=total deaths). By subtraction, we estimated total disabilities (total serious harms–total deaths=total disabilities).

Uncertainty estimates were calculated using a probabilistic sampling approach based on Monte Carlo simulations[24] (full statistical R V.4.2.2 code is provided in online supplemental file 2). In this manuscript, many ranges are denoted 'probabilistic plausible ranges' (PPRs), rather than 95% CIs. This is because they rely on some diagnostic error rates (n=5 cancers) that use literature-derived (and expert-validated) plausible ranges (PRs) rather than statistically derived 95% CIs, reflecting uncertainty beyond mere sampling error[8] (online supplemental file 1-A5). We used PRs for the top five cancers because different studies defined diagnostic delays of different lengths—defining shorter delays as errors created an upper PR bound, while defining longer delays as errors created a lower PR bound.[8]

Sensitivity analyses and validity checks

We used five separate approaches to assess the robustness of our final results: (1) sensitivity analyses using different data assumptions ((a) one-way analyses to assess the impact of uncertainty in model parameters by using the lower and higher uncertainty bounds rather than the point estimate and (b) the impact of analysing disease incidence for vascular events and infections using only principal NIS diagnoses) (online supplemental file 1-B1,B2); (2) assessing the risk of misestimating deaths by undercounting (incident cases resulting in prehospital death) or overcounting (patients admitted more than once in a given year, yet who could only die once) (online supplemental file 1-B3,B4); (3) comparison with independent hospital and autopsy estimates (online supplemental file 1-C1,C2); (4) triangulation of data derived from studies of diagnostic errors and harms across clinical settings (inpatient, emergency department, ambulatory clinics)

(online supplemental file 1-C3) and (5) an iterative process of expert review by 24 clinical domain experts (following the same method used in our prior publication to validate estimates of error and harm rates),[8] which served as a final check on the face validity of our disease-specific incidence and total harm estimates.

Statistical analysis and reporting

We used sample sizes, totals, means, medians, 95% CIs, IQRs and PPRs to describe populations and outcomes, as appropriate. NIS analysis was conducted using the PROC SURVEYMEANS procedure in SAS V.9.3 (Cary, North Carolina, USA). All other statistical calculations were performed using R V.4.2.2 (Vienna, Austria). This manuscript follows Enhancing the QUAlity and Transparency Of Health Research (Strengthening the Reporting of Observational Studies in Epidemiology)[25] reporting guidelines for observational studies.

Role of the funding source

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

RESULTS

Quality of data sources for error and harm rates

Error and harm rates were published previously.[3 8] For 14 of 15 diseases (besides arterial thromboembolism, where we aggregated four retrospective case series), condition-specific diagnostic error rates were derived from high-quality clinical literature. This included clinical studies with strong designs (large prospective clinical trials or studies using population-based sampling or registries) or meta-analyses of high-quality clinical studies. For condition-specific diagnostic error rates, there were 47 source studies (vascular events (n=28), infections (n=10), cancers (n=9)) representing 942 916 patients (median study sample n=397 (IQR 176–1914); median per-disease sample n=2343 (IQR 398–10 351)). For disease-agnostic harm rates, there were five source studies representing 1216 diagnostic errors and 374 serious harms.[8] Each study operationalised definitions slightly differently (eg, nature of diagnostic reference standard), but all definitions for errors/harms were consistent with published definitions described in the 'Methods' section.

US population-based incidence of vascular events, infections, and cancers

The total NIS sample from 2012 to 2014 included 21.5 M hospitalisations (for all conditions, not just vascular events or infections), representing a weighted national estimate of 107.4 M total discharges (mean annual 35.8 M). In 2014, the sample was taken from 4411 different hospitals across 45 states (representing ~80% of hospitals and 90% of states in the USA). The mean weighted annual number of incident vascular events was 6.0 M (95% CI 5.9 to 6.0). Patients had a median age of 67.5 years (IQR 57.2–78.2, range 0–90); 44.8% were female and 70.0% were non-Hispanic white. The mean weighted annual number of incident infections was 6.2 M (95% CI 6.1 to 6.3). Patients had a median age of 63.7 years (IQR 52.8–79.8, range 0–90); 51.3% were female and 68.6% were non-Hispanic white. The number of incident cancer cases in 2014 was 1.5 M. Patients had a median age of just over 65 years (<20, 0.9%; 20–49, 11.8%; 50–64, 33.2%; 65–74, 28.5%; ≥75, 250.7%); 50.7% were female and 80.0% were non-Hispanic white. The estimated total annual incidence of all 'Big Three' diseases was 13.7 M (43.5% vascular events, 45.2% infections, 11.3% cancers) (table 1).

Overall incidence of diagnostic errors and serious harms

Table 1 shows annual estimated disease incidence, diagnostic errors, and serious misdiagnosis related harms by disease and by category (and denotes whether uncertainty for each parameter is represented by CI, PR, or PPR). Serious misdiagnosis-related harms are summarized in Figures 1 and 2. Across the 'Big Three' categories, there were 1.51M (PPR 1.12-1.89) missed diagnoses and 603,000 (PPR 454,000-776,000) serious harms; mean diagnostic error and serious harm rates per true disease case for any 'Big Three' disease (including 'other' subcategories) were 11.1% and 4.4%, respectively. The 15 individually analyzed 'Big Three' diseases together accounted for 403,000 serious harms (50.7% of the grand total); mean diagnostic error and serious harm rates per true disease case for the 15 specific diseases (excluding 'other' subcategories of the 'Big Three') were 11.1% and 6.1%, respectively. Among these, five conditions linked to the largest numbers of serious harms (stroke, sepsis, pneumonia, venous thromboembolism, and lung cancer) together accounted for 308,000 serious harms (38.7% of the grand total). Across all dangerous diseases (including non 'Big Three'), the grand total estimate was 2.59M (PPR 1.92-3.23) missed diagnoses and 795,000 (PPR 598,000-1,023,000) serious harms (broken down as 371,000 total deaths and 424,000 total disabilities).

Sensitivity analyses and validity checks

The population-level serious harm totals were most sensitive to harm rates for the highest-incidence infections ('other' infections, sepsis, pneumonia) and stroke, but even if each of these harm rates were placed at the lower plausible bound of harms for that specific disease, the grand total of serious harms across all diseases would still be over 500 000 (online supplemental file 1-B1). Using only principal diagnosis NIS codes, which assumes a lower disease incidence and reduces any residual risks of double counting, gave lower estimates by about 30% (grand totals 1.78 M missed diagnoses and 549 000 serious harms (online

Table 1 Annual US incidence of dangerous diseases, diagnostic errors and serious misdiagnosis-related harms

'Big Three' disease	Disease incidence* n, in thousands (95% CI)	Diagnostic error rate* % (95% CI, PR, PPR†)	Diagnostic errors n, in thousands (PPR)	Serious misdiagnosis-related harm rate* % (PPR)	Serious harms n, in thousands (PPR)
Vascular					
Stroke	952 (937 to 967)	17.5% (95% CI 9.5 to 27.3)	166 (90–260)	9.8% (5.3–15.5)	94 (51–148)
Venous thromboembolism	320 (315 to 324)	20.4% (95% CI 17.0 to 23.9)	65 (54–77)	10.9% (8.9–13.1)	35 (28–42)
Arterial thromboembolism	173 (170 to 176)	23.9% (95% CI 18.9 to 29.5)	41 (33–51)	12.7% (9.9–16.0)	22 (17–28)
Aortic aneurysm and dissection	96 (93 to 99)	35.6% (95% CI 21.0 to 51.7)	34 (20–50)	22.1% (13.0–32.5)	21 (12–31)
Myocardial infarction	1242 (1219 to 1266)	1.5% (95% CI 1.0 to 2.2)	19 (13–27)	0.8% (0.5–1.2)	10 (7–15)
Top five vascular events subtotal	2783 (2754 to 2811)	11.7% (PPR 8.8–15.1)	326 (245–421)	6.5% (4.9–8.5)	182 (136–237)
Other vascular events	3173 (3131 to 3215)	11.7% (PPR 8.8–15.1)‡	372 (279–480)	1.4% (1.1–1.9)	46 (34–59)
Total vascular events	5956 (5905 to 6006)	11.7% (PPR 8.8–15.1)	697 (524–900)	3.8% (2.9–5.0)	228 (170–296)
Infection					
Sepsis	1345 (1325 to 1365)	9.9% (95% CI 2.8 to 20.6)	134 (38–278)	5.9% (1.7–12.3)	79 (23–165)
Pneumonia	1469 (1452 to 1486)	9.5% (95% CI 2.3 to 14.3)	140 (34–210)	4.6% (1.1–7.0)	68 (16–103)
Meningitis and encephalitis	47 (46 to 48)	25.6% (95% CI 20.8 to 30.8)	12 (10–15)	14.7% (11.8–18.1)	7 (6–9)
Spinal abscess	14 (13 to 14)	62.1% (95% CI 54.6 to 69.2)	8 (7–9)	36.7% (31.5–42.2)	5 (4–6)
Endocarditis	34 (33 to 35)	25.5% (95% CI 21.7 to 29.6)	9 (7–10)	13.8% (11.5–16.4)	5 (4–6)
Top five infections subtotal	2909 (2882 to 2936)	10.4% (PPR 4.6–14.9)	303 (133–435)	5.6% (2.5–8.3)	164 (73–242)
Other infections	3286 (3249 to 3323)	10.4% (PPR 4.6–14.9)‡	342 (151–491)	3.3% (1.4–4.7)	107 (47–154)
Total infections	6195 (6150 to 6241)	10.4% (PPR 4.6–14.9)	645 (284–925)	4.4% (1.9–6.4)	271 (120–395)
Cancer					
Lung cancer	224§	22.5% (PR 11.3–37.8)	50 (25–85)	14.2% (7.1–24.1)	32 (16–54)
Breast cancer	235§	8.9% (PR 8.5–26.3)	21 (20–62)	4.5% (4.2–13.4)	11 (10–31)
Colorectal cancer	137§	9.6% (PR 8.4–47.7)	13 (12–65)	5.6% (4.9–28.1)	8 (7–38)
Melanoma	76§	13.6% (PR 6.8–25.0)	10 (5–19)	5.7% (2.8–10.6)	4 (2–8)
Prostate cancer	233§	2.4% (PR 1.7–13.8)	6 (4–32)	1.3% (0.9–7.4)	3 (2–17)
Top five cancers subtotal	905§	11.1% (PPR 10.1–20.9)	100 (92–189)	6.3% (5.6–12.0)	57 (51–108)
Other cancers	640§	11.1% (PPR 10.1–20.9)‡	71 (65–134)	7.4% (6.7–14.2)	47 (43–91)
Total cancers	1545§	11.1% (PPR 10.1–20.9)	171 (156–323)	6.8% (6.1–12.8)	105 (94–198)
Additional totals					
Total big three (top five only)	6597 (6558 to 6636)	11.1% (PPR 8.3–13.8)	729 (549–913)	6.1% (4.6–7.8)	403 (305–511)
Total big three (top five+other)	13 697 (13 628 to 13 765)	11.1% (PPR 8.2–13.8)	1514 (1122–1889)	4.4% (3.3–5.7)	603 (454–776)
Grand total¶	N/A¶	N/A¶	2588 (1918–3230)	N/A¶	795 (598–1023)

*Disease incidence as measured here is an estimate of total 'true disease' cases (rather than only 'correctly diagnosed' cases). Diagnostic error and serious misdiagnosis-related harm rates were published previously[3 8] (reference #3: stroke, venous thromboembolism, aortic aneurysm and dissection, myocardial infarction, sepsis; reference #8: all other individual diseases). These rates derive from studies of 'true disease' cases. The 'diagnostic error rate' and 'serious misdiagnosis-related harm rate' are both given with respect to the overall dangerous disease incidence. For example, for stroke (shown in the first content row of table 1): (a) diagnostic errors are derived as ~952 000 (column #2)×17.5% (column #3)=~166 000 (column #4); (b) serious misdiagnosis-related harms are derived as ~952 000 (column #2)×9.8% (column #5)=~94 000 (column #6).

†Shown are either 95% CIs, PRs or PPRs. True statistical 95% CIs were used when data allowed their calculation without expert input. PRs were used 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; PRs were thus defined and determined based partially on input from relevant domain experts, as described in the 'Methods' section, so reflect more than just sampling-related variability. PPRs derive from Monte Carlo analysis, which included a mix of diagnostic error rates that used 95% CIs and those that used PRs. Because simulations used some PRs (n=5 cancers), all Monte Carlo results are reported as PPRs.

‡Because we could not estimate error rates for the residual, unnamed non-top five 'other' diseases within each 'Big Three' category, we used the mean error rate for the top five diseases (eg, for unnamed 'other' vascular events, we used the mean diagnostic error rate for stroke, venous thromboembolism, arterial thromboembolism, aortic aneurysm and dissection and myocardial infarction). We used disease incidence-weighted means (eg, the error rate for myocardial infarction had proportionally more impact on the final mean than the error rate for aortic aneurysm and dissection, because there are ~13-fold more incident cases of myocardial infarction). PPRs derive from Monte Carlo analysis.

§Because North American Association of Central Cancer Registries and American Cancer Society treat estimates as a complete census of cases (ie, no sampling-related uncertainty), no 95% CIs are represented.

¶The 'Grand Total' is calculated from the 'Big Three' to all dangerous diseases causing serious misdiagnosis-related harms, based on the proportion of errors (58.5%) and serious harms (75.8%) attributable to the 'Big Three' in previously published clinical literature.[8] Thus, no estimates are provided for 'disease incidence', 'diagnostic error rate' or 'serious harm rate' columns. PPRs derive from Monte Carlo analysis.

PPR, probabilistic plausible range; PR, plausible range.

supplemental file 1-B2)). The impact of methodological assumptions on undercounting (online supplemental file 1-B3) and overcounting (online supplemental file 1-B4) were both estimated at <8% and likely offsetting.

Validity checks assessed current results based on similarity to (or coherence with) values derived independently using setting-specific (eg, hospital-based) medical literature. Estimated misdiagnosis-attributable death rates were 14.1% (n=~371 000 of 2.6 M US deaths in 2014) for the primary analysis and 9.8% (n=~256 000 of 2.6 M US deaths in 2014) for the principal-only analysis (online supplemental file 1-C1). By comparison, the literature-derived rate of misdiagnosis-attributable deaths based on hospital autopsies (8.4%, 95% CI

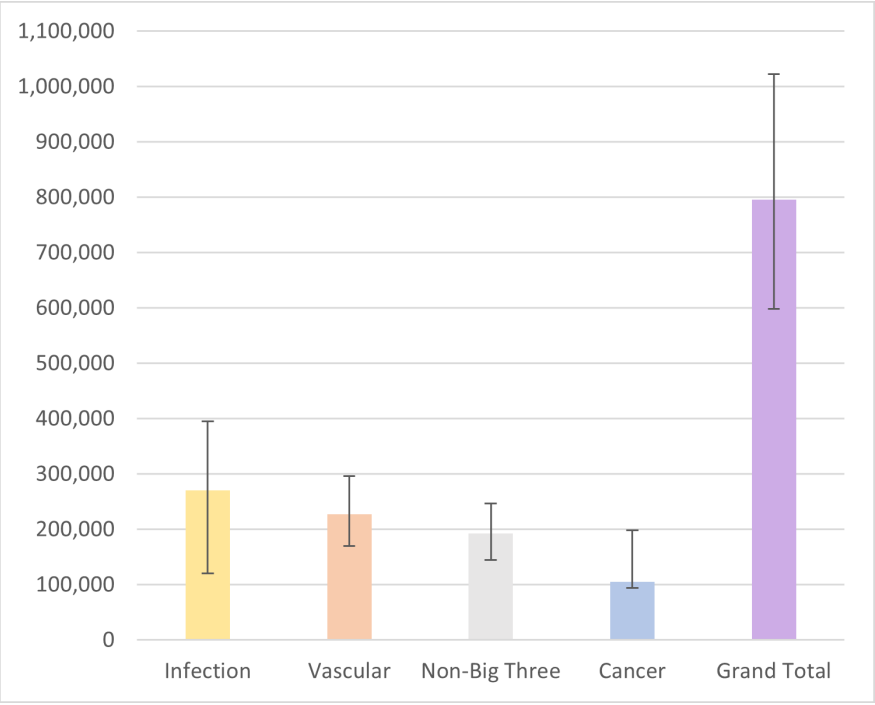

Figure 1 Annual population incidence of serious misdiagnosis-related harms from vascular events, infections, cancers and all non-'Big Three' others. The estimated grand total annual US incidence for serious harms (combining 'Big Three' harms with other non-'Big Three' harms) is 795 000 (probabilistic plausible range (PPR) 598 000–1 023 000). Whiskers denote PPRs from the Monte Carlo analysis.

5.2 to 13.1) and inpatient diagnostic adverse events (~7.4%) were lower, as expected (online supplemental file 1-C1,C2). Our disease-based estimate of total serious misdiagnosis-related harms (across clinical settings) of ~795 000 (PPR 598 000–1 023 000) was comparable to independent literature-derived values using a setting-based (rather than disease-based) approach, which assessed ~855 000 (490 000-1 659 000) serious misdiagnosis-related harms (online supplemental file 1-C3). Estimates of inpatient misdiagnosis-related deaths derived from our disease-based approach (~105 000) fall within the uncertainty bounds of those derived independently from previously published medical literature on hospital autopsies (~82 000 (51 000–128 000)) and hospital-based adverse events

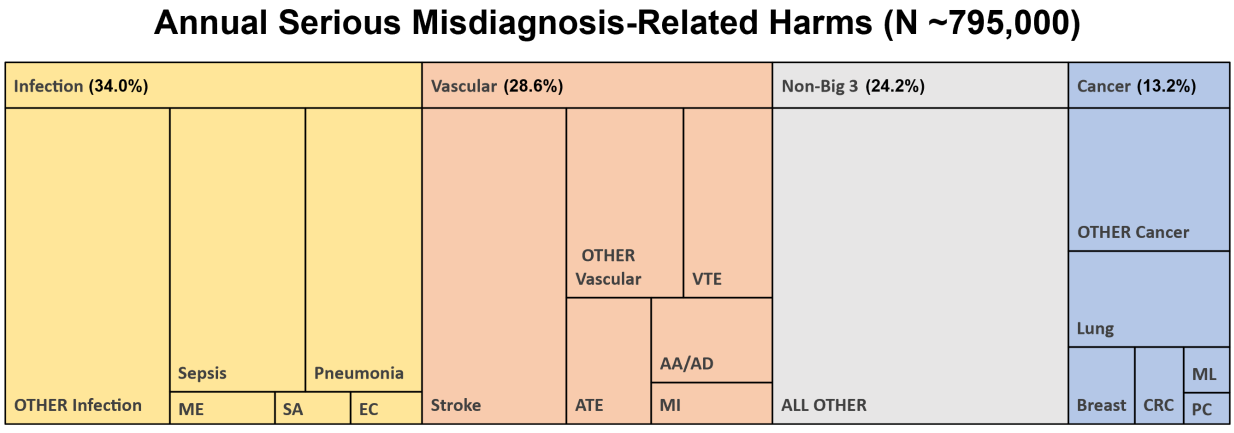

Figure 2 Fraction of serious misdiagnosis-related harms in the USA attributed to the top diseases by category. The treemap diagram proportionally represents hierarchical categories and specific diseases causing serious harms when the diagnosis is incorrect. As we reported previously, based solely on clinical studies, 'Big Three' diseases account for 75.8% of all serious harms.[9] The current analysis shows these are broken down as 34.0% infections, 28.6% vascular events and 13.2% cancers. Taken together, the top five vascular events, infections and cancers account for 50.7% of all serious harms; the five most frequently harmful conditions across 'Big Three' categories account for 38.7% of all serious harms. AA/AD, aortic aneurysm/aortic dissection; ATE, arterial thromboembolism; CRC, colorectal cancer; EC, endocarditis; ME, meningitis/encephalitis; MI, myocardial infarction; ML, melanoma; PC, prostate cancer; SA, spinal abscess; VTE, venous thromboembolism.

(~72 000 (51 000–113 000)) (online supplemental file 1-C3). Per-visit serious harm rates by either method were estimated at 0.08% (online supplemental file 1-C3). After iterative review and feedback (described previously),[8] final estimates for disease-specific incidence, error/harm rates and total serious harms were deemed face valid by 24 clinical domain experts.

DISCUSSION

This manuscript provides the first robust, national annual US estimate for serious misdiagnosis-related harms (nearly 800 000 combined deaths (~371 000) or permanent disabilities (~424 000)) across care settings (ambulatory clinic, emergency department and inpatient). Even with the most conservative assumptions about disease incidence or disease-specific harms, we estimated the number affected to be over 500 000. The number of affected patients is large, and this makes diagnostic error a pressing public health concern. Our results also suggest that meaningful progress could be made by addressing just a few dangerous diseases that are relatively common—reducing diagnostic errors by ~50% for the 15 named dangerous diseases could potentially prevent ~200 000 serious misdiagnosis-related harms while reducing diagnostic errors by ~50% for the five most harmful diseases (stroke, sepsis, pneumonia, venous thromboembolism and lung cancer) could prevent ~150 000.

Sensitivity analyses and validity checks show serious misdiagnosis-related harm results are robust. The impact of methods-induced undercounting and overcounting were relatively small and likely cancel one another. The credibility of our current estimate is bolstered by convergent construct validity with two alternative methods of estimation using the rate of misdiagnosis-attributable deaths based on hospital autopsies and inpatient diagnostic adverse events. Care setting-based estimates using independent, disease-agnostic data from two large systematic reviews (inpatient[1] and emergency department[3]) also corroborate our findings.

Our results suggest that diagnostic error is probably the single largest source of deaths across all care settings (~371 000) linked to medical error. This number may exceed estimated deaths from all other patient safety concerns combined, regardless of which prior estimate of total deaths due to medical error (range 12 500–250 000[26]) is considered. This seems plausible because prior estimates systematically undercount diagnostic errors and diagnostic errors more often cause serious harms than other errors.[27]

How many misdiagnosis-associated disabilities or deaths are preventable and how much (or little) longevity might potentially be reclaimed for affected patients is uncertain. Preventability is inconsistently judged by different raters, and some remain sceptical that error prevention can meaningfully increase longevity with a good quality of life.[28] Nevertheless, there are numerous anecdotes of otherwise healthy young patients in whom a half-century or more of quality life years are likely to have been saved through prompt diagnosis.[29] For some of the most harmful diseases in our list, correct initial diagnosis has been associated with substantial reductions in morbidity or mortality (eg, ischaemic stroke (~fivefold),[3] aneurysmal subarachnoid haemorrhage (~fivefold),[30] ruptured abdominal aortic aneurysm (~twofold)).[31] Finally, large variation in diagnostic error and harm rates across demographic groups, diseases, clinical settings and individual institutions point to strong prospects of preventability for at least some harms.[3 32]

Although the study estimated total diagnostic errors (2.59 M), this reflects only errors in patients with dangerous diseases, not all diagnostic errors. Total annual diagnostic errors in the USA likely number in the tens of millions, but the total is likely highly contingent on the threshold for defining a diagnostic error.[8] This is different, however, than serious harms (death and permanent disability), which are more objectively defined, so less subject to this particular type of methodological heterogeneity.[8]

The large absolute numbers of patients harmed should not be mistaken for an inordinately high per-incident case or per-visit risk. According to these results, a patient with a life-threatening or limb-threatening disease has a ~11% chance of being missed; because of the substantial risk of harm when a dangerous disease is missed, that same patient also has a ~4% overall chance of dying or becoming permanently disabled pursuant to a misdiagnosis. Admittedly, both are higher than what medical experts generally think of as an 'acceptable' miss rate for dangerous diseases (eg, <0.5%–1%).[33–35] However, given over 1 billion healthcare visits per year in the USA,[8] a patient visiting a doctor for any reason (ie, who may or may not have a dangerous underlying disease) likely has a <0.1% chance of suffering serious misdiagnosis-related harms. Thus, patients should not panic or lose faith in the healthcare system.

Although the present study focused on US-based estimates, some of our disease-specific error rates were based on data from other high-income countries outside the USA,[3 8] and there is good reason to believe that diagnostic errors and misdiagnosis-related harms represent a global problem. There is meta-analytic evidence that hospital-based diagnostic error and harm rates are comparable across North America and Europe, but higher in other countries that were studied.[1] Measured error and harm rates in primary care[4 6 36] and emergency departments[3] are similar in the USA, the UK and Western Europe. In 2015, Organisation for Economic Co-operation and Development (OECD) nations averaged 6.5 doctor consultations per person year[37] and had ~1.3 billion persons[38]—if per-visit serious harm rates are comparable to the USA, this would translate to roughly 7 M serious misdiagnosis-related harms in OECD nations (including the USA). Less

is known about the scope and nature of diagnostic errors in low-income and middle-income nations. However, access to basic diagnostic testing resources are very limited in many low-income and middle-income countries,[39 40] and diagnostic delays for life-threatening diseases can be substantial,[41–43] so the global burden for ~7.9 billion persons is likely several-fold higher.

Disease distributions for serious misdiagnosis-related harms differ across clinical settings and age groups. Missed vascular events and infections dominate in hospitals and emergency departments, while missed cancers likely dominate in primary care.[3 9] In adult care, vascular events are typical, while in paediatric care, infections are typical.[3] Thus, diseases that should be the focus of interventions to improve diagnostic performance would ideally be tailored to the specific clinical context.

This study focused on missed diagnoses (false negatives) of dangerous diseases. While it is desirable to prevent false negatives, practical realities may constrain our ability to do so. Implications for improving diagnosis must consider these results in the broader diagnostic context which includes overuse of diagnostic tests, false positive (mis-)diagnoses, incidental findings and overdiagnosis,[10] because these are also associated with substantial harms[12 44 45] and increased healthcare costs.[46] Reducing missed diagnoses by increasing sensitivity at the expense of specificity (ie, trading false negatives for false positives by shifting clinical decision thresholds around ordering tests or interpreting test results) should not be considered 'improving diagnosis'.[47 48] Instead, diagnostic innovations that increase both sensitivity and specificity at a given test threshold are needed,[47] as recently shown in a pilot tele-consult programme for dizziness and stroke in the emergency department.[49] Economic modelling may be an important means to estimate the full future impact of solutions designed to improve diagnosis, before they are implemented.[47]

Limitations

Our approach relies on literature-derived estimates being roughly representative of US national diagnostic error and serious harm rates, which cannot be directly verified. Although some estimates based on older studies might not generalise to current practice, limited available evidence suggests diagnostic errors are either stable or rising over time in the USA.[3 50] Population-based incidence estimates for vascular events and infections using the NIS are based on administrative codes that could not be independently clinically verified by our team, but annual disease-specific incidence values were deemed face valid by relevant specialists. Our approach is limited by drawing together data from several sources, each with its own uncertainty, so our final estimates are necessarily less precise than would be desirable. This estimate does not account for the sometimes profound effects of non-disabling suffering due to diagnostic delays of non-lethal illnesses, including prolonged diagnostic odysseys,[51] chronic side effects and risks of treatments administered for diseases patients do not actually have (false positives)[52 53] and the substantial health effects and economic consequences of overtesting[12 44] and overdiagnosis.[45] Nevertheless, our national extrapolations are based on current best evidence regarding error/harm rates, triangulate well with data from other sources and are face valid to disease-specific domain experts.

CONCLUSIONS

Across clinical settings (ambulatory clinics, emergency department and inpatient), we estimate that nearly 800 000 Americans die or are permanently disabled by diagnostic error each year, making it the single largest source of serious harms from medical mistakes. We believe this is the best estimate currently possible, and, in an area of patient safety where estimates vary widely, results presented here offer an important scientific advance for the field. Although not all these harms are necessarily preventable, our findings add urgency to what the US National Academy of Medicine has already labelled a 'moral, professional, and public health imperative'. Policymakers have recently taken notice,[54] but diagnostic error-related research still remains substantially underfunded relative to its public health impact[48]—to make progress, this must change. Research and quality improvement programmes should include a strong focus on prompt diagnosis of vascular events, infections and cancers, with an emphasis on the top 15 dangerous diseases identified in this study, which together likely account for half of all serious misdiagnosis-related harms. Prospective, interventional studies are needed to confirm the real-world preventability of these harms.

Author affiliations
[1]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
[2]Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA
[3]Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
[4]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
[5]Department of Patient Safety, The Risk Management Foundation of the Harvard Medical Institutions Inc, Boston, Massachusetts, USA
[6]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA
[7]Department of Oncology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
[8]Candello, The Risk Management Foundation of the Harvard Medical Institutions Inc, Boston, Massachusetts, USA
[9]Department of Risk Management & Analytics, Coverys, Boston, Massachusetts, USA

Correction notice This article has been corrected since it was first published online. The following section has been removed from the funding statement: (EPC VI (TOPIC ID 503-4262).

Twitter David E Newman-Toker @DNewmanToker

Contributors DEN-T (guarantor): I accept full responsibility for the finished work and the conduct of the study, had access to the data, and controlled the decision to publish. I declare that I designed the study; had primary oversight over the data analysis; designed the figures; authored the primary

manuscript draft and all major revisions and that I have seen and approved the final version. I served as an unpaid member of the Board of Directors of the Society to Improve Diagnosis in Medicine, and as its President (2018–2020). I serve as a medico-legal consultant for both plaintiff and defence in cases related to diagnostic error. I have no other relevant conflicts of interest. NN: I declare that I assisted in study design and conduct; edited the manuscript for scientific content and that I have seen and approved the final version. I have no conflicts of interest. ACS: I declare that I assisted in study design and conduct; edited the manuscript for scientific content and that I have seen and approved the final version. I have no conflicts of interest. CWY-M: 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. ASST: 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. GDC: 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. ZW: I declare that I designed the statistical analysis, including Monte Carlo simulations to create probabilistic plausible range estimates; edited the manuscript for scientific content and that I have seen and approved the final version. I have no conflicts of interest. YZ: I declare that I assisted in the design and conduct of the statistical analysis, including Monte Carlo simulations to create probabilistic plausible range estimates; edited the manuscript for scientific content and that I have seen and approved the final version. I have no conflicts of interest. MF: 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. AH: 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. DS: I declare that I assisted in study design and conduct; edited the manuscript for scientific content and that I have seen and approved the final version. I previously served as an unpaid member of the Board of Directors of the Society to Improve Diagnosis in Medicine.

Funding This study was funded by Society to Improve Diagnosis in Medicine, Agency for Healthcare Research and Quality (R01 HS 27614, R18 HS 029350) and Armstrong Institute Center for Diagnostic Excellence at the Johns Hopkins University School of Medicine.

Competing interests DEN-T has a career focus and conducts research related to diagnostic errors, including in patients with dizziness and stroke. He serves as the principal investigator for multiple grants and contracts on these topics. DEN-T is a former volunteer President and member of the Board of Directors of the Society to Improve Diagnosis in Medicine. Johns Hopkins has been loaned research equipment (video-oculography (VOG) systems) by two companies for use in DEN-T's research; one of these companies has also provided funding for DEN-T's research on diagnostic algorithm development related to dizziness, inner ear diseases and stroke. DEN-T has no other financial interest in these or any other companies. DEN-T is an inventor on a provisional patent (US PCT/US2020/070304) for smartphone-based stroke diagnosis in patients with dizziness. He gives frequent academic lectures on these topics and occasionally serves as a medico-legal consultant for both plaintiff and defence in cases related to dizziness, stroke and diagnostic error. DS is also a former volunteer 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 organisations that could inappropriately influence (bias) their work.

Patient consent for publication Not applicable.

Ethics approval No human subjects participated in this study, and no institutional review board approval was needed.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data on disease incidence used for the study are all publicly available; these public-use datasets and accompanying standard data dictionaries may be found at the URL locations cited in the references list. Additional details regarding sources and methods for diagnostic error and harm rate calculations may be found in three prior publications (PMID: 31535832, 32412440, 36574484), including their associated appendices and online supplemental materials.

ORCID iD

David E Newman-Toker http://orcid.org/0000-0003-2789-4115

REFERENCES

1 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.
2 Wilson RM, Michel P, Olsen S, *et al*. Patient safety in developing countries: retrospective estimation of scale and nature of harm to patients in hospital. *BMJ* 2012;344:e832.
3 Newman-Toker DE, Peterson SM, Badihian S, *et al*. Diagnostic errors in the emergency Department: a systematic review. comparative effectiveness review No.258. (prepared by the Johns Hopkins University evidence-based practice center under contract No.75Q80120D00003.) AHRQ publication No.22(23)-Ehc043. Rockville, MD Agency for Healthcare Research and Quality; 2022.
4 Avery AJ, Sheehan C, Bell B, *et al*. Incidence, nature and causes of avoidable significant harm in primary care in England: retrospective case NOTE review. *BMJ Qual Saf* 2021;30:961–76.
5 Slawomirski L, Auraaen A, Klazinga N. The Economics of patient safety in primary and ambulatory care: flying blind. Paris, 2018.
6 Singh H, Giardina TD, Meyer AND, *et al*. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013;173:418–25.
7 Improving diagnosis in Healthcare [Institute of Medicine]. 2015. Available: http://www.nationalacademies.org/hmd/Reports/2015/Improving-Diagnosis-in-Healthcare.aspx [Accessed 20 May 2023].
8 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.
9 Newman-Toker DE, Schaffer AC, Yu-Moe CW, *et al*. Serious misdiagnosis-related harms in malpractice claims: the "big three" - vascular events, infections, and cancers. *Diagnosis (Berl)* 2019;6:227–40.
10 Newman-Toker DE. A unified conceptual model for diagnostic errors: underdiagnosis, overdiagnosis, and misdiagnosis. *Diagnosis (Berl)* 2014;1:43–8.

11 Segal JB, Nassery N, Chang H-Y, *et al*. An index for measuring overuse of health care resources with Medicare claims. *Med Care* 2015;53:230–6.
12 Smith-Bindman R. Use of advanced imaging tests and the not-so-incidental harms of incidental findings. *JAMA Intern Med* 2018;178:227.
13 Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102:605–13.
14 NAIC malpractice claims, final compilation. Brookfield, WI National Association of Insurance Commissioners; 1980. Available: https://www.naic.org/documents/prod_serv_special_med_lb.pdf [Accessed 22 Apr 2019].
15 Guideline for implementation of medical professional liability closed claim reporting (GDL-1077). National Association of Insurance Commissioners; 2010. Available: http://www.naic.org/store/free/GDL-1077.pdf [Accessed 22 Apr 2019].
16 Introduction to the HCUP national inpatient sample [Agency for Healthcare Research and Quality]. 2014. Available: https://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2014.jsp#table3app1 [Accessed 20 May 2023].
17 NAACCR Fast Stats. North American Association of central cancer registries (NAACCR). Available: https://faststats.naaccr.org [Accessed 21 Sep 2018].
18 Annual cancer facts and figures [American Cancer Society]. 2014. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2014/cancer-facts-and-figures-2014.pdf [Accessed 20 May 2023].
19 McCormick N, Lacaille D, Bhole V, *et al*. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One* 2014;9:e92286.
20 ICD-10-CM official guidelines for coding and reporting FY 2019 [Centers for Medicare & Medicaid Services]. 2019. Available: https://www.cms.gov/Medicare/Coding/ICD10/Downloads/2019-ICD10-Coding-Guidelines-.pdf [Accessed 20 May 2023].
21 Tirschwell DL, Longstreth WT. Validating administrative data in stroke research. *Stroke* 2002;33:2465–70.
22 Producing national HCUP estimates - accessible version [Agency for Healthcare Research and Quality]. 2015. Available: https://www.hcup-us.ahrq.gov/tech_assist/nationalestimates/508_course/508course_2018.jsp [Accessed 20 May 2023].
23 Zwaan L, de Bruijne M, Wagner C, *et al*. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med* 2010;170:1015–21.
24 Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970;57:97–109.
25 von Elm E, Altman DG, Egger M, *et al*. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573.
26 Rodwin BA, Bilan VP, Merchant NB, *et al*. Rate of preventable mortality in hospitalized patients: a systematic review and meta-analysis. *J Gen Intern Med* 2020;35:2099–106.
27 Saber Tehrani AS, Lee H, Mathews SC, *et al*. 25-year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National practitioner data bank. *BMJ Qual Saf* 2013;22:672–80.
28 Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA* 2001;286:415–20.
29 Society to Improve Diagnosis in Medicine (SIDM) - patient stories. Society to Improve Diagnosis in Medicine. Available: https://www.improvediagnosis.org/stories [Accessed 20 May 2023].
30 Kowalski RG, Claassen J, Kreiter KT, *et al*. Initial misdiagnosis and outcome after subarachnoid hemorrhage. *JAMA* 2004;291:866–9.
31 Smidfelt K, Nordanstig J, Davidsson A, *et al*. Misdiagnosis of ruptured abdominal aortic aneurysms is common and is associated with increased mortality. *J Vasc Surg* 2021;73:476–83.
32 Chang T-P, Bery AK, Wang Z, *et al*. Stroke hospitalization after misdiagnosis of "benign dizziness" is lower in specialty care than general practice: a population-based cohort analysis of missed stroke using SPADE methods. *Diagnosis (Berl)* 2021;9:96–106.
33 Pines JM, Szyld D. Risk tolerance for the exclusion of potentially life-threatening diseases in the ED. *Am J Emerg Med* 2007;25:540–4.
34 Newman-Toker DE, Edlow JA. High-stakes diagnostic decision rules for serious disorders: the Ottawa subarachnoid hemorrhage rule. *JAMA* 2013;310:1237–9.
35 Kene MV, Ballard DW, Vinson DR, *et al*. Emergency physician attitudes, preferences, and risk tolerance for stroke as a potential cause of dizziness symptoms. *West J Emerg Med* 2015;16:768–76.
36 Cheraghi-Sohi S, Holland F, Singh H, *et al*. Incidence, origins and Avoidable harm of missed opportunities in diagnosis: longitudinal patient record review in 21 English general practices. *BMJ Qual Saf* 2021;30:977–85.
37 Health at a glance 2015: OECD indicators. Paris OECD Publishing; 2015.
38 Population, total - OECD members [The World Bank]. Available: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=OE [Accessed 20 May 2023].
39 Yadav H, Shah D, Sayed S, *et al*. Availability of essential diagnostics in ten low-income and middle-income countries: results from national health facility surveys. *Lancet Glob Health* 2021;9:e1553–60.
40 Chimatiro GL, Rhoda AJ. Scoping review of acute stroke care management and rehabilitation in low and middle-income countries. *BMC Health Serv Res* 2019;19:789.
41 Evenson KR, Foraker RE, Morris DL, *et al*. A comprehensive review of prehospital and in-hospital delay times in acute stroke care. *Int J Stroke* 2009;4:187–99.
42 Storla DG, Yimer S, Bjune GA. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC Public Health* 2008;8:15.
43 Brand NR, Qu LG, Chao A, *et al*. Delays and barriers to cancer care in low- and middle-income countries: a systematic review. *Oncologist* 2019;24:e1371–80.
44 Smith-Bindman R, Miglioretti DL, Johnson E, *et al*. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. *JAMA* 2012;307:2400–9.
45 Black WC. Overdiagnosis: an underrecognized cause of confusion and harm in cancer screening. *J Natl Cancer Inst* 2000;92:1280–2.
46 Ong MS, Mandl KD. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. *Health Aff (Millwood)* 2015;34:576–83.
47 Newman-Toker DE, McDonald KM, Meltzer DO. How much diagnostic safety can we afford, and how should we

decide? A health economics perspective. *BMJ Qual Saf* 2013;22 Suppl 2:ii11–20.
48 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. Available: https://www.improvediagnosis.org/roadmap [Accessed 20 May 2023].
49 Gold D, Peterson S, McClenney A, *et al*. Diagnostic impact of a device-enabled remote "TELE-dizzy" consultation service [abstract]. Diagnostic Error in Medicine, 12th Annual Conference; Washington, DC, 2019
50 Waxman DA, Kanzaria HK, Schriger DL. Unrecognized cardiovascular emergencies among medicare patients. *JAMA Intern Med* 2018;178:477–84.
51 Kliegman RM, Bordini BJ, Basel D, *et al*. How doctors think: common diagnostic errors in clinical judgment-lessons from an undiagnosed and rare disease program. *Pediatr Clin North Am* 2017;64:1–15.
52 Solomon AJ, Bourdette DN, Cross AH, *et al*. The contemporary spectrum of multiple sclerosis misdiagnosis: a multicenter study. *Neurology* 2016;87:1393–9.
53 Campisano J. My turn: repeal of Obamacare means repeal of protections for patients like me. Available: https://www.azcentral.com/story/opinion/op-ed/2017/02/19/repeal-obamacare-pre-existing-condition/97746562 [Accessed 20 May 2023].
54 Improving diagnosis in medicine acts of 2019 and 2020 [Congress.gov]. 2019 and 2020. Available: https://www.congress.gov/bill/116th-congress/house-bill/5014/text and https://www.congress.gov/bill/116th-congress/senate-bill/5002 [Accessed 20 May 2023].

Supplemental File #1 for ***Burden of Serious Harms from Diagnostic Error in the USA***
by Newman-Toker, et al. BMJ Quality & Safety, 2023 (doi:10.1136/bmjqs-2021-014130)

Enclosed below are the main Supplementary Materials (Supplement 1) to the peer-reviewed, scientific journal publication entitled ***Burden of Serious Harms from Diagnostic Error in the USA*** *(2023).* Some of the methods descriptions (particularly in Supplement 1, section A5 about Monte Carlo analysis) are very similar to methods descriptions from the related, previously published manuscripts from earlier project phases. This is unavoidable since the statistical methods for the current manuscript were the same. An ancillary appendix (Supplement 2) provides the full statistical code for the Monte Carlo analysis.

Table of Contents for Supplement 1

Abbreviations Not Necessarily Defined in the Text

N/n – number
M – million
B – billion
CI – confidence interval
PR – plausible range
PPR – probabilistic plausible range

A) Methodological Details

A1. Study Methods – Rationale for Overall Architecture

Seeking valid estimates of disability and death from diagnostic error is important for public policy, yet methodologically challenging.[1] Diagnostic errors will necessarily be more frequent than diagnostic adverse events (of any severity), which, in turn, will be more frequent than serious misdiagnosis-related harms (i.e., permanent disability or death). Key concerns in estimation of serious harms[1] include (1) proper quantitative synthesis of the literature on error and harm rates, rather than extrapolations based on single studies conducted in non-representative settings; (2) judgments about attributable harms (i.e., the extent to which diagnostic errors result in serious harms or are potentially confounded by comorbidity); (3) judgments about preventability, including whether prevention will result in meaningful gains in healthy life years; (4) methodological risks of undercounting or overcounting serious harms, including double counting of deaths in patients who suffer more than one error; and (5) applying error and harm rate estimates to the appropriate population at risk and conducting methodologically robust statistical analyses to account for uncertainty associated with relatively low frequency events (i.e., serious harms).

Taking a disease-specific (rather than disease-agnostic) approach to measurement helps address many of these methodological challenges. It is easier to more rigorously and precisely measure diagnostic errors, harms, or preventability in disease-specific than disease-agnostic fashion since research studies that cut across diseases cannot incorporate rigorous reference standards for diagnostic accuracy or error for every possible condition. Synthesis of multiple studies via systematic review with meta-analysis is also more straightforward since disease-specific studies are more homogeneous in their disease and error definitions. Finally, a consistent finding from the literature on diagnostic error, whether derived from malpractice claims or clinical practice, is that vascular events, infections, and cancers (together known as the "Big Three" dangerous-disease categories) are responsible for three-quarters of serious misdiagnosis-related harms.[2,3] This permits extrapolation from these specific disease categories to all diseases.

As noted in the main manuscript Methods section, the overall study was designed to estimate the total annual burden (incidence) of serious harms from diagnostic error in the US. It was conducted in three study phases (Figure S1): (1) identify top diseases misdiagnosed that cause serious misdiagnosis-related harms (from a large, nationally representative malpractice data set previously coded for claim type, outcome/harm severity, and disease then comparing the proportion of "Big Three" diseases to clinical practice-based [non-claims] studies)[2]; (2) find disease-specific diagnostic error and harm rates for top harm-causing diseases (from clinical literature and vetted by experts)[4,5]; (3a) measure annual population incidence for each disease (from public use, nationally representative data sets) and (3b) combine error and harm rates with incidence to estimate total annual US incidence of serious diagnostic errors and harms. The first two study phases were published previously[2,4]; the final phase (3a, 3b) is included here.

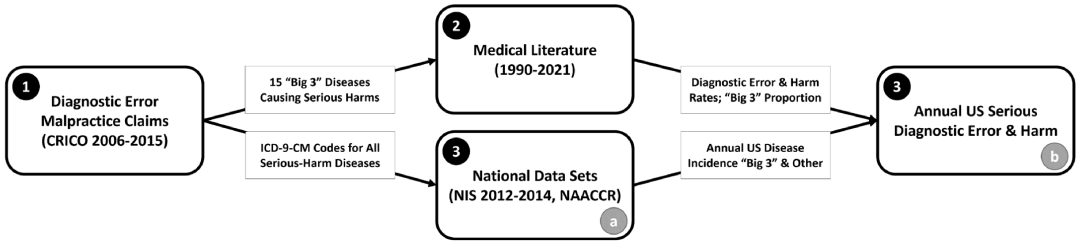

Figure S1. Overview of scientific methods for estimating total diagnostic errors and serious misdiagnosis-related harms. *Abbreviations: CRICO – Controlled Risk Insurance Company, Ltd.; ICD-9-CM – International Classification of Diseases 9th Revision, Clinical Modification; NAACCR – North American Association of Central Cancer Registries; NIS – National Inpatient Sample.*

A2. Role of Malpractice Claims in Estimation and Independence of Grand Total Harm Estimates

Some readers may wonder how it can be that our study began with a first phase that used malpractice claims data, yet the final estimates are somehow independent? As noted in the main manuscript Methods section, our scientific approach was constructed such that the final grand total estimates for errors and harms in the US are based on clinical literature and US population incidence, not malpractice claims. This is because (a) no error or harm rates were taken from claims-based studies, (b) the extrapolation from specific "Big Three" disease estimates to the grand total were based on the proportion of "Big Three" diseases causing errors and harms from clinical studies, and (c) any impact of having used malpractice claims to construct the original disease list or weights are mathematically unrelated to the grand totals.

Why were malpractice data used in the first place? Malpractice data were used to construct the initial list of diseases likely to be responsible for the greatest numbers of total misdiagnosis-related harms. We needed a starter list of such diseases in order to be able to take the disease-based approach (A1 above).

If malpractice data were good enough to make the disease list, why not just stop there? First, it is known that many medical errors never lead to malpractice claims, so it is hard to extrapolate even from a representative, national claims database to a true national estimate. In the Harvard Medical Practice Study, which compared negligent medical errors to malpractice claims, the chance that an injury caused by medical negligence would result in litigation was just 1.5% (95% CI 0-3.2).[6] Second, malpractice claims data are known to represent a biased sample. Some forms of bias in malpractice claims are well known, while others may be hidden. The most well-known bias in malpractice claims data is towards higher-severity harms. This is not necessarily a problem for estimating serious harms[7] (as in this study), but there may also be maldistributions of claims (i.e., non-representativeness) based on other factors as well. For example, myocardial infarction is probably overrepresented in claims relative to stroke as a cause of misdiagnosis-related harms—disease incidence is similar, diagnostic error rates are ~10-fold higher for stroke, and disabling neurologic injuries result in the highest claims payouts (so are more likely to spark a claim), yet numbers of claims are only ~1.5-fold higher for stroke; this could be because legal "standard of care" expectations for accuracy of heart attack diagnoses are higher than those for stroke.[5]

Were malpractice data used for any other purpose in constructing the estimates? Yes, as noted in the main manuscript Methods section, misdiagnosis-related harm rates were derived by combining high-quality data on disease-agnostic (non-disease specific) harms per diagnostic error from well-respected clinical studies then applying disease-specific harm-severity weights from malpractice claims.[4] A disease-agnostic approach was required because there were not a sufficient number of disease-specific studies examining attributable harm rates. We weighted the disease-agnostic, per-diagnostic-error serious harm rate for each disease to get a more realistic estimate of harms (e.g., aortic dissection is more likely lethal than pneumonia when initially missed, so assigning the same risk of serious harms *per diagnostic error* for each of the two diseases would have been inappropriate). For each disease, we multiplied the disease-specific, clinical literature-based diagnostic error rate ***by*** the clinical literature-based disease-agnostic per-error harm rate ***by*** a disease-specific, claims-based harm-severity weight. This weight was based on the disease-specific proportion of malpractice cases resulting in serious vs. non-serious harms (e.g., higher weight for aortic dissection than pneumonia). The weighting procedure was also used to prevent overcounting of harms from "other" (non-top 5) diseases. Full statistical details of this approach can be found in our prior publication's Supplementary materials (*Supplement A2, Requirements R1 and R4*[4]).

If that is true, then how can the final total estimates be claims-independent? The final results are independent of malpractice claims because we mathematically "forced" the proportion of errors and serious harms attributable to all combined Big Three diseases to be equal to the known attributable fractions found in the clinical literature (*see our prior publication's Supplement A2, Requirements R2 and*

R3[4]). This is described, in brief, in the Methods section of the main manuscript, "…"Big Three" results were used to calculate a grand total (including non-"Big Three" dangerous diseases) using the clinical proportion of diagnostic errors (58.5%) and serious harms (75.8%) attributable to "Big Three" diseases.[2,3] These proportions derive exclusively from research studies based in clinical practice (i.e., not malpractice claims studies) (*see Table 3 from our prior citation*[2]). Mathematically, the grand total of diagnostic errors was calculated by dividing the "Big Three" total number of diagnostic errors by 0.585. Similarly, the grand total of serious misdiagnosis-related harms was calculated by dividing the "Big Three" total number of serious misdiagnosis-related harms by 0.758." This forces independence from claims.

Then what are the implications of malpractice-claims based intermediate steps for the results? There are two main potential impacts of these claims-based steps. First, it is possible that the lower-ranked "top 5" diseases might be over-ranked (e.g., it is possible that the unnamed 6th-ranked disease categorized in the "other" subcategory in one of the "Big Three" categories might actually be the *real* 5th-ranked disease in that category). For example, the 6th-ranked disease in the infection category in malpractice claims was appendicitis. There were more than twice as many claims for endocarditis as appendicitis, which is why we searched out data on error rates for endocarditis rather than appendicitis. However, if malpractice claims were somehow biased towards endocarditis or away from appendicitis, it is still potentially conceivable that appendicitis might outrank endocarditis as a cause of misdiagnosis-related harms in clinical practice, since appendicitis has more than twice the real-world incidence of endocarditis. However, this is unlikely, because endocarditis is initially missed an estimated ~26% of the time[4] and appendicitis is initially missed no more than ~5% of the time, more than compensating for the higher incidence of appendicitis.[5] Note that, in our final analysis, appendicitis is still accounted for in the "other" infections subcategory (so it has not gone uncounted). Second (and related), it is possible that the relative proportion of "other" (non-top 5) diseases are underrepresented relative to the top 5.

In summary, serious harms estimated for individual diseases named (or unnamed) in Table 1 of the main manuscript are potentially impacted by unknown biases that could be present in malpractice claims. This could impact disease-specific rankings or the proportion of "other" (non-top 5) harms. However, the grand total harm estimates are mathematically fully independent of malpractice claims.

A3. Double Check of HCUP CCS Code Level Groupings Prior to NIS Incidence Analysis

Prior to NIS analysis of disease incidence in this third phase of the project, we performed a final cross-check at the code level using the HCUP CCS Level 3 groupings for vascular diseases and infections derived from the claims analysis.[2] The code lists were reviewed and any ICD codes unrelated to new, acute events (e.g., 438 "late effects of cerebrovascular disease") were removed prior to NIS analysis. We also reviewed all codes in the CRICO CBS data set to address issues of coding migration over time and reduce the risk that any specific codes might be missed because of sampling error in CRICO data during the years of analysis. From the wider code list, we found 18 related codes that belonged in the top 5 groupings (e.g., 433.0 "occlusion and stenosis of the basilar artery" and 435.3 "vertebrobasilar artery syndrome" for stroke) and added these before conducting the final NIS analysis. We did not consider Level 3 codes present in the parent HCUP CCS classification but *not* found in the malpractice claims data, to avoid any risk of overcounting non-life-or-limb-threatening diseases unlikely to cause harms. NIS analysis was run at both the ICD-9-CM code level as well as rolled up by disease and category to both (a) ensure sensibility and coherence and (b) identify any coding errors or gaps before being finalized.

A4. NIS Sampling & Weighting Procedures to Derive Nationally Representative Estimates

We followed standard procedures for NIS data to derive nationally representative estimates, which use pre-specified discharge weights to convert an unweighted sample of hospital discharges into a weighted,

nationally representative sample.[8] The result is a weighted estimate for both disease incidence and patient demographics. Each year in the US there are roughly 36M inpatient hospitalizations[9] at more than 6,000 hospitals.[10] For each year since 2012, NIS has sampled more than 7M hospital discharge records from more than 4,000 acute-care hospitals (excluding long-term acute care hospitals). The discharge weights are calculated by NIS data curators by first stratifying the NIS hospitals on the same variables that were used for creating the sample. These variables are hospital Census division, urban/rural location, teaching status, bed size, and ownership. A weight is then calculated for each stratum, by dividing the number of universe discharges (i.e., all discharges) in that stratum, obtained from HCUP and American Hospital Association data, by the number of NIS discharges (i.e., sampled discharges) in the stratum. Discharge weights are assigned to each sampled discharge by NIS data curators and are stored in the NIS data set for use in constructing nationally representative estimates. When discharge weights are applied to the unweighted NIS data, the result is an estimate of the number of discharges for the entire universe (i.e., an estimate of all acute care hospitalizations in the US).

A5. Monte Carlo Analysis to Determine Probabilistic Plausible Ranges (PPRs) *(reported previously[4])*

The main outcome measures were estimates of total US annual diagnostic errors and serious misdiagnosis-related harms. Annual incidence from NIS and NAACCR were multiplied by literature-derived estimates of disease-specific and category diagnostic error and harm rates,[4] an approach analogous to "minimal modeling" methods in cost-effectiveness or value-of-information analysis.[11]

To obtain the variability of these combined estimates, we used a probabilistic sampling approach based on Monte Carlo simulations[12] (Supplement 2). These simulations produce statistically valid 95% CIs that account for variability in both number and sample sizes for each disease. In the current manuscript, most of these uncertainty estimates are denoted as "probabilistic plausible ranges" (PPRs), rather than 95% CIs. This is because they rely, in part, on diagnostic error rates that utilize literature-derived (and expert-validated) plausible ranges (n=5 cancers) rather than statistically derived 95% CIs as their uncertainty range, reflecting some uncertainty beyond mere sampling error.[4] Specifically, experts felt that for the top five cancers, PRs should be wider than the statistical 95% CIs. For each cancer, this was because different studies defined diagnostic delays of different lengths—defining shorter delays as errors created an upper PR bound, while defining longer delays created a lower PR bound.[4] As part of the same Monte Carlo simulations, we also calculated PPRs around error and harm point estimates for the "other" (non-top 5) subcategories and combined categories (e.g., top 5 vascular events, total "Big Three," grand totals).

For the Monte Carlo analysis, skew-normal distributions were used to approximate the distributions of disease incidence rate, diagnostic error rate, and serious misdiagnosis-related harm rate, separately for each quantity. The location parameter of the skew-normal was set to be the point estimate of the corresponding rate. The scale and skewness parameters were determined such that the lower and upper bounds of 95% CI of the resulting skew-normal distribution coincided with each of the 95% CI or probabilistic range bounds of the corresponding rate. Due to the extreme skewness of plausible ranges for some diagnostic error rates,[4] all approximations were performed on logit-transformed distributions.[13] Monte Carlo samples were drawn independently from the resulting distributions. The population affected by the diagnostic error and the subsequent misdiagnosis-related harm were calculated for each Monte Carlo replica. The PPRs were given by the 2.5% and 97.5% quantiles based on 10,000,000 simulations. The large number of simulations was used to ensure tail probability and reduce Monte Carlo error due to the very skewed sampling distributions. Rates and other parameters have been published previously,[4] although diagnostic error rate estimates for stroke, myocardial infarction, venous thromboembolism, aortic aneurysm and dissection, and sepsis were updated to reflect the most robust estimates available from a systematic review with meta-analysis that was conducted by members of the authorship team.[5]

B) Estimated Impact of Methodological Choices & Assumptions on Results

B1. Sensitivity Analysis of Errors & Harms (Impact of Uncertainty in Model Parameters)

We conducted a one-way sensitivity analysis to assess the impact of parameter uncertainty on the final point estimates (Figure S2). Harm results were most sensitive to parameters for common infections. The three highest leverage parameters for potentially overestimating serious harms were the (1) other infection harm rate, (2) sepsis harm rate, and (3) pneumonia harm rate. However, even if each of these harm rates (and the one for stroke [4th for overestimates]) were placed at the lower plausible bound of harms assessed in this one-way sensitivity analysis, the point estimate of serious harms would still be over 500,000.

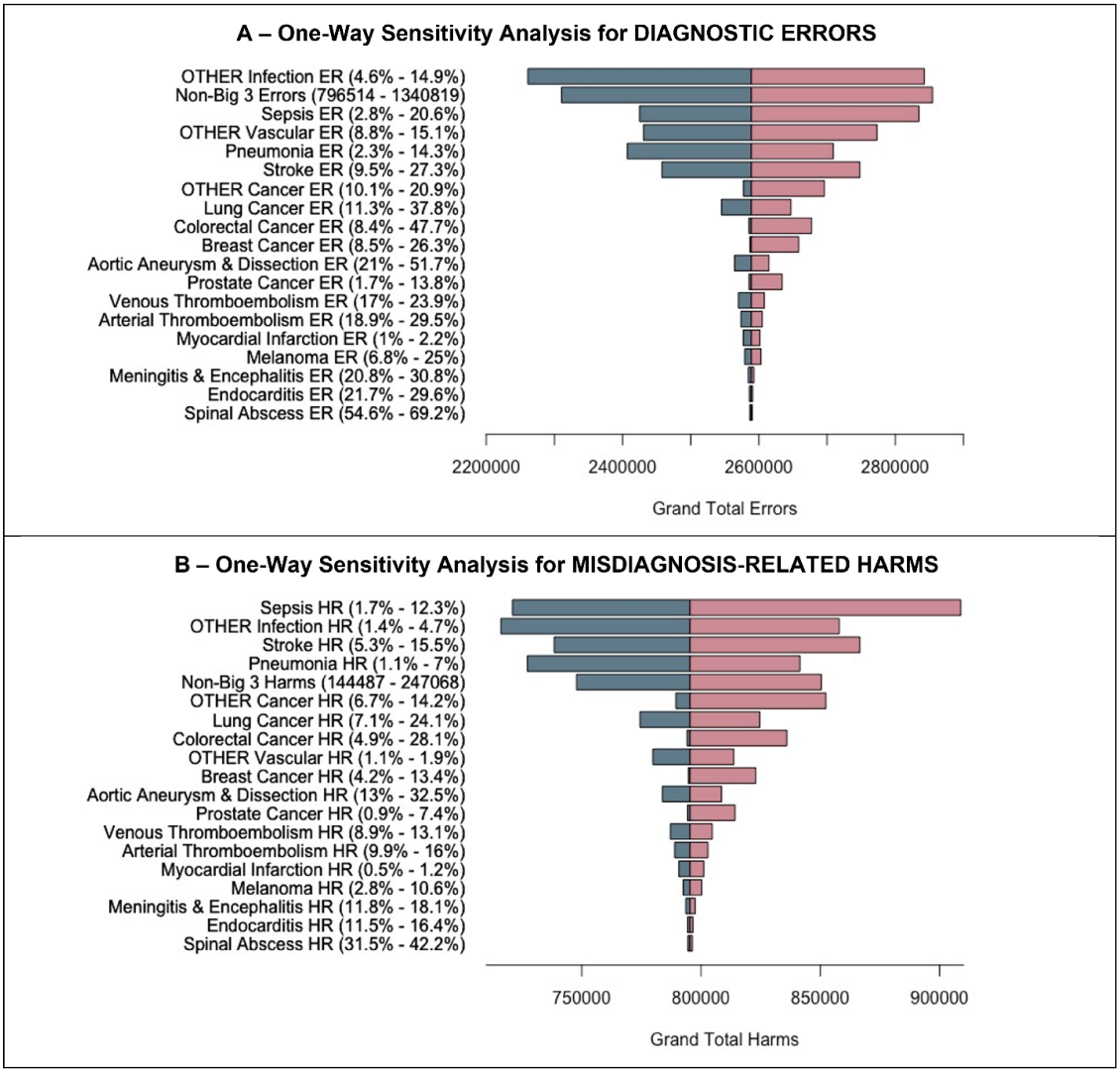

Figure S2. Sensitivity analysis of model parameters on diagnostic errors and serious harms. This one-way (as opposed to multi-way, probabilistic sensitivity analysis) varies one parameter value at a time while holding all other model parameter values constant. For each parameter, its value is ranged between the lower and upper uncertainty bounds (i.e., confidence interval or plausible range [*shown*]) and the impact of uncertainty on the total is illustrated, with highest impact parameters shown at the top and lowest at the bottom, yielding a "tornado" appearance. The

point estimate value used for each parameter is fixed as the "midline" of the tornado. The impact on the final total of using the lower bound parameter value is shown in ***blue to the left*** (reflecting possible ***overestimation*** in the point estimate). Conversely, the impact on the final total of using the upper bound parameter value is shown in ***red to the right*** (reflecting possible ***underestimation*** in the point estimate). Panel A shows a tornado diagram for diagnostic errors. The three parameters with the greatest potential for overestimation of errors were the (1) other infection error rate, (2) non-"Big Three" error rate, and (3) pneumonia error rate. Panel B shows a tornado diagram for serious harms. The three parameters with the greatest potential for overestimation of harms were the (1) other infection harm rate, (2) sepsis harm rate, and (3) pneumonia harm rate. *Abbreviations: ER – error rate; HR – harm rate.*

B2. Sensitivity Analysis of Errors & Harms (Impact of Using Only Principal Diagnosis)

We estimated disease incidence for vascular events and infections from HCUP data from the NIS. In the primary analysis, we counted discharge (or in-hospital death) diagnoses coded in either the principal or first-listed secondary diagnosis positions, as these two diagnoses are often of equal, competing weight.[14]

We conducted a sensitivity analysis using only first-position NIS diagnosis codes to assess the impact of this methodological decision on the final results. Primary vs. sensitivity results are shown in Table S1. Overall estimates of errors and harms were about 30% lower when using only principal diagnosis codes.

Table S1. Comparison of diagnostic errors and harms in the primary analysis vs. principal-only analysis*

Category	Primary Analysis N in thousands (PPR)	Principal Diagnosis-Only N in thousands (PPR)
Big Three Total Diagnostic Errors	1,514 (1,122-1,889)	1,044 (852-1,365)
Big Three Total Serious Harms	603 (454-776)	416 (344-550)
Grand Total Diagnostic Errors	2,588 (1,919-3,230)	1,785 (1,457-2,335)
Grand Total Serious Harms	795 (598-1,023)	549 (454-725)

* The primary analysis counted NIS diagnosis codes in either the principal or first-listed secondary positions. The sensitivity analysis counted NIS diagnosis codes in only the principal position (so are necessarily lower).

B3. Estimated Impact of Undercounting (Conservative Assumption re: Out-of-Hospital Deaths)

The conservative assumption was made that incident cases of dangerous (life or limb-threatening) vascular events and infections in the US would eventually involve a hospitalization, even if the patient was initially misdiagnosed in an ambulatory care setting. Outpatient (e.g., primary care, emergency department) visit diagnoses were not included separately in the disease incidence calculations because they would risk inflating disease incidence estimates through double counting. For example, had "myocardial infarction" cases that were correctly diagnosed in outpatient care (and then later confirmed as an inpatient) been included in the analysis, the same incident cases would be counted twice.

This methodological decision, chosen to avoid overcounting, does risk some degree of undercounting; this is principally via patients seen in outpatient settings (e.g., primary care or emergency department) who are misdiagnosed and then die in the community of their underlying illness, without ever reaching the hospital as an inpatient (i.e., out-of-hospital deaths following a missed diagnosis). Reliable data are not available to directly measure out-of-hospital deaths across all conditions (because death certificate data may be unreliable[15]). However, we can estimate the incidence of such deaths using pre-hospital death data, which are available for some of the more common dangerous diseases. For myocardial infarction, the proportion of all true cases that result in pre-hospital death was estimated in a rigorous population-based study from Germany to be 13.6%.[16] For stroke, we can estimate the rate by combining data on pre-hospital stroke deaths with data on stroke hospitalizations. In 2014 (the reference year for our study) there were 980 stroke hospitalizations per 100,000 among Medicare beneficiaries (who are predominantly patients aged 65 and older) in the state of New York.[17] With a 2014 New York state population aged 65

and older of 2,898,094,[18] the total number of stroke hospitalizations for those aged 65 and older in 2014 was 28,401. Since ~38% of stroke hospitalizations in 2014 occurred in patients below age 65,[19] the total number of New York state stroke hospitalizations in 2014 was approximately ~45,800. In 2012-2014, the crude rate of out-of-hospital stroke deaths in New York state was 11.6 per 100,000.[20] With a 2014 New York state total population of ~19,750,000,[18] the total number of pre-hospital stroke deaths was about ~2,300. Therefore, the proportion of strokes that resulted in death without making it to the hospital was about 4.8% of all strokes. The heart attack and stroke estimates cohere well with what is known about the natural history of these two diseases—harms from myocardial infarction are disproportionately deaths, while harms from stroke are disproportionately disabilities. Accordingly, it is sensible that a higher proportion of myocardial infarctions would result in pre-hospital deaths. If we take a disease incidence-weighted average of these two numbers (13.6% for heart attack and 4.8% for stroke), the estimated proportion of undercounted incident cases is ~10%.

If we postulate a similar overall rate (~10% out-of-hospital deaths) for the remaining vascular events and infections, and if 100% of these out-of-hospital death cases were deemed "misdiagnosed," our overall estimate of serious misdiagnosis-related harms would be under-counted by approximately 8% (since this particular problem is not likely to impact incident cancers, which are measured by different means of estimation). However, many such cases involve patients who die without ever having had the opportunity to be misdiagnosed because they never reached medical attention (e.g., previously asymptomatic sudden cardiac death from myocardial infarction[21] or strokes in which premonitory transient ischemic attack symptoms do not prompt the individual to seek attention[22]) (Figure S3). ***Thus, 8% represents a likely upper bound on undercounting and the true value is probably less than 5%***.

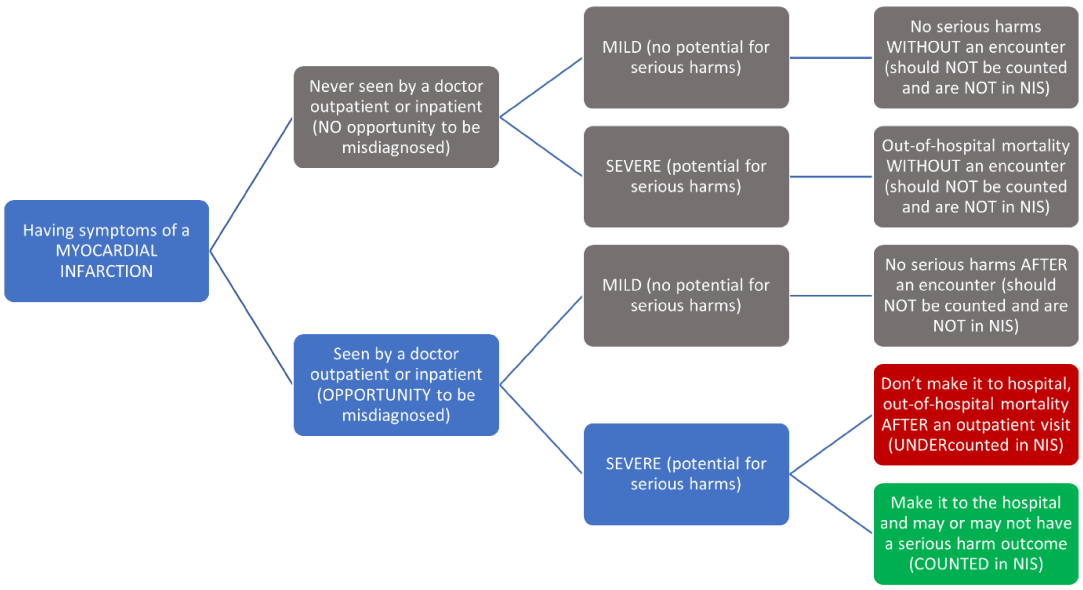

Figure S3. Conceptual relationship between disease, care-seeking behavior, and risk of undercounting. The figure illustrates how a specific subset of undercounted cases not in NIS data (shown in red) might impact the overall estimates of serious misdiagnosis-related harms; other undercounted cases would not (shown in gray). As described in the text, it is estimated that the impact of these undercounted cases is likely less than 5% of the total.
Abbreviation: NIS – National Inpatient Sample

B4. Estimated Impact of Overcounting (Based on Patients with More than One Hospitalization)

There is some possibility that our method of using inpatient hospitalizations to measure dangerous disease incidence might lead to overcounting. NIS data track hospitalizations, not patients, so some patients could have been admitted more than once (e.g., admitted for a myocardial infarction and later a stroke in the same year). Although a single person could suffer permanent disability in more than one way, one patient cannot die twice, so this could theoretically lead to overestimates of deaths using our method.

It is not possible to estimate the impact of such potential overcounting directly using NIS data, but we can estimate the potential extent of the problem by combining NIS with other data sources, such as the National Health Interview Survey (NHIS). Using NHIS, the Centers for Disease Control and Prevention (CDC) reports that, in 2014, ~17.5M (corresponding to ~76% of patients hospitalized at least once that year) were hospitalized only once, ~3.2M (corresponding to ~14% of patients hospitalized at least once that year) were hospitalized twice, and ~2.3M (corresponding to ~10% of patients hospitalized at least once that year) were hospitalized three or more times.[23] With ~35.4M total hospitalizations that year in NIS,[24] that means ~50% of hospitalizations involved "repeat visitors." Using NIS data from our current analysis (average for 2012-2014), 34% of hospitalizations were for vascular events or infections (i.e., 12.1M of 35.8M total hospitalizations). The clinical proportion of serious misdiagnosis-related harms represented by deaths is 46.7%.[25,26] Thus, the potential impact on overcounting deaths from missed vascular events and infections is ~8% (i.e., ~50% x ~34% x ~47% = ~8%). However, patients who did, in fact, die of a misdiagnosis from one of these illnesses could not have been counted again past their death date (i.e., they could not have been a "repeat visitor"), ***so the true value is likely to be lower (e.g., <5%)***.

C) Additional Validity Arguments

C1. Comparison with Independent Estimates from Diagnostic Errors in Hospital Autopsies

To gauge the plausibility of our overall serious misdiagnosis-related harms estimate (~795,000), we can derive the misdiagnosis-associated mortality in our data and compare it to that found in hospital autopsy data. We estimate total deaths from our current study using the previously published proportion of harms representing deaths across inpatient and outpatient settings—46.7%[25,26] (~795,000 x 46.7% = ~371,000). We can do the same for the principal-only analysis (~549,000 x 46.7% = ~256,000). Table S2 shows total expected US deaths and all serious harms (death plus permanent disability) in 2014 (our year of analysis), depending on the hypothesized proportion of deaths associated with diagnostic error.

Table S2. Anticipated US deaths and serious harms due to diagnostic error depending on hypothesized risk

Hypothetical Proportion of Deaths Associated with Diagnostic Error	Potential Misdiagnosis-Related Deaths in the US in 2014*	Corresponding Misdiagnosis-Related Serious Harms in 2014†
5%	131,321	281,027
10%	262,642	562,053
15%	393,963	843,080
20%	525,284	1,124,107

***** According to the Centers for Disease Control and Prevention (CDC), there were 2,626,418 (2.6M) US deaths in 2014,[27] which is the reference year chosen for our manuscript's analysis of serious misdiagnosis-related harms.
† Calculated using the clinical proportion of serious misdiagnosis-related harms represented by deaths (46.7%).[25,26]

The primary analysis estimate (~371,000 deaths [of ~2.6M deaths]) would represent a 14.1% (10.6-18.2) overall misdiagnosis-associated mortality nationally. The principal-only analysis (~256,000 deaths [of ~2.6M deaths]) would represent a 9.8% (8.1% to 12.9%) overall misdiagnosis-associated mortality

nationally. Either rate is higher than estimates from studies of hospital autopsies that consider epoch of diagnosis and adjust for bias from submaximal autopsy rates. A large meta-analysis of hospital autopsy studies projected that a modern US hospital which autopsied 100% of in-hospital deaths would find 8.4% (95% CI 5.2-13.1) suffered a major diagnostic error, half considered Class I (deaths directly attributed to the diagnostic error) and half Class II (diagnostic errors that would have changed clinical management and could have altered the patient's clinical course).[28]

However, it is expected that the population-based proportion of misdiagnosis-related deaths would be higher than that found in hospital autopsies. One reason is that hospital autopsies consider diagnostic errors not yet recognized at the time of death, but usually not errors occurring pre-hospitalization when a prompt intervention might have been lifesaving. For example, consider a patient with a new, abrupt-onset headache who is sent home from a primary care clinic as "migraine." If the patient were to return a week later to the emergency department in a coma, they might be promptly diagnosed with aneurysmal subarachnoid hemorrhage and admitted to the intensive care unit. Were they to die, a hospital autopsy would then indicate that no diagnostic error had occurred *in the hospital*. The risk of death from brain aneurysm is increased nearly 5-fold after an initial misdiagnosis, and misdiagnosis disproportionately occurs in outpatient clinics with isolated headache clinical presentations.[29,30] Given highly effective treatments for brain aneurysm and the knowledge that prognosis post-operatively is almost entirely tied to clinical severity at the time of surgery,[29] this case should clearly count as a potentially preventable death due to diagnostic error, but would be considered a correct ante-mortem diagnosis in hospital autopsy data.

Another reason is that the proportion of deaths associated with diagnostic error/delay is probably higher for out-of-hospital than in-hospital deaths. Estimated diagnostic error rates in primary care (2.4% per visit [n=5,126/212,165][31,32]) exceed those in hospitals (0.7%[33]). This makes sense, since (a) hospitalization tends to occur relatively late in the natural course of illness, when a patient has become sick enough to merit inpatient care, and often after the underlying cause for their symptoms is more obvious, and (b) hospitalized patients undergo more intensive diagnostic testing and monitoring than ambulatory patients. As a result, it would not be surprising if outpatient deaths were more often pursuant to diagnostic errors. Since >80% of healthcare visits occur in non-ED ambulatory care[34,35] and >65% of all US deaths occur outside the hospital,[36] an overall misdiagnosis-associated mortality of 9.8% to 14.1% seems plausible.

How many of these misdiagnosis-associated deaths are preventable and how much (or little) longevity might potentially be reclaimed for affected patients is uncertain.[37] Nevertheless, individual cases of otherwise healthy young patients who die from treatable causes that were misdiagnosed make it clear that this could be a half-century or more in years of quality life lost for a given patient.[38-40] The same is true for lifelong disability in young patients after missed opportunities to promptly treat disabling diseases.[41-43] For some of the most harmful diseases in our list, correct initial diagnosis has been associated with clear and substantial reductions in morbidity or mortality (e.g., ischemic stroke [~5-fold],[5] aneurysmal subarachnoid hemorrhage [~5-fold],[29] and ruptured abdominal aortic aneurysm [~2-fold][44]).

C2. Comparison with Independent Estimates from Diagnostic Adverse Events in Hospitals

We can also gauge the plausibility of our serious harm results in light of diagnostic adverse event data from inpatient hospital stays. Gunderson et al. recently published a systematic review of hospital-based studies of diagnostic adverse events (n=22), two of these US-based.[33] They estimated a pooled hospital misdiagnosis-related harm rate (counting any harm severity) of 0.7% (95% CI 0.5-1.1) with high levels of heterogeneity (I^2=95%, p<0.001) (overall range across studies 0.1-2.7).[33] Most of these studies did not report specific diseases missed, but eight did (n=136 cases). Authors listed 70 diseases or categories with at least two instances (Table 2[33]). Among these 70, 78.6% were attributed to "Big Three" diseases (this distribution is very similar to the attributable % used in our current population-based study [75.8%[2]]).

If applied to US-based hospitals, they estimated ~250,000 patients harmed annually from diagnostic error. They were unable to assess harm severity based on the available literature. In the well-designed 2010 Dutch study by Zwaan et al.,[25] which measured a similar rate of hospital-based diagnostic adverse events (0.4%), they found that, of diagnostic adverse events, 29.1% resulted in death and ~25.6% (estimated from their Figure 1) resulted in disability at discharge, for an overall rate of ~54.7% serious harms.[25]

Using NIS 2014 US hospitalizations (~35.4M),[24] this translates to ~135,000 (~97K-213K) (uncertainty estimated using 95% CI from Gunderson[33]) serious misdiagnosis-related harms in US hospitals. Relative to our 2014 primary analysis estimate of ~795,000 suffering death or permanent disability, this suggests that ~17% (~12-27) of serious misdiagnosis-related harms occur among inpatients and ~83% (~73-88) among outpatients. Although in 2014 only 2.7% of the roughly 1.3B US healthcare visits were inpatient hospitalizations,[4] severity of illness and diagnostic error adverse events are both higher than outpatient.[45] To help gauge this effect, the proportion of high-severity misdiagnosis-related harms linked to inpatient care in malpractice claims is ~28% (i.e., >10-fold over-representation relative to visit proportion).[2] It is reasonable to expect that inpatient malpractice claims for diagnostic adverse events would be artificially over-represented in claims relative to clinical care proportions, since, relative to outpatient care, outcome severity is higher[45] (a known predictor of legal action[7]) and a "paper trail" of documentation to establish a legal action is more readily available (another likely predictor[2]). Therefore, the ~28% represents an "upper bound," of sorts, on the inpatient-attributable serious harms fraction. Thus, an estimate that ~17% (~12-27) of US serious misdiagnosis-related harms occur in inpatient settings seems quite plausible.

Using these numbers, we can also estimate that the total annual hospital-based deaths from diagnostic error in the US in 2014 would be ~35.4M (NIS 2014[24]) x 0.7% (95% CI 0.5-1.1) (Gunderson et al.[33]) x 29.1% (Zwaan et al.[25]) = ~72,000 (~51K-113K). This value is squarely within the range projected by Leape, Berwick, and Bates (i.e., 40,000-80,000 hospital deaths per year)[46] derived by multiplying total hospital deaths by the rate of hospital autopsy-determined diagnostic errors. The ~72,000 misdiagnosis-related hospital deaths estimate is also squarely in the range of what is expected based on a rigorous systematic review of hospital autopsies by Shojania et al.[28] They calculated the combined Goldman Class I/II diagnostic error rate for an average, modern, US-based hospital that autopsied 100% of its deaths—8.4% (95% CI 5.2-13.1).[28] According to the CDC, there were 2,626,418 (2.6M) US deaths in 2014,[27] of which 37.3% were hospital-based,[47] for a total of ~980,000 hospital deaths. The ~72,000 would therefore correspond to a 7.4% (n=~72,000/~980,000) misdiagnosis-attributable fraction of hospital deaths.

The resulting estimates comparing inpatient-only harms to those across settings are shown in Table S3.

Table S3. US deaths and serious harms due to diagnostic error in 2014 comparing inpatient to all settings

Misdiagnosis-Related Harms	Inpatient Only (Prior Studies[25,33])	Across Settings (Current Study)
Total Serious (Death + Disability)	~135,000 (~97K-213K)*	~795,000 (~598K-1,023K)†
Deaths Only	~72,000 (~51K-113K)*	~371,000 (~279K-478K)‡
Disability (*calculated difference*)	~63,000 (~46K-100K)	~424,000 (~319-545)

* Uncertainty accounted for using 95% CI from Gunderson[33] plus serious harms or death % from Zwaan.[25]
† Uncertainty accounted for using PPR from primary analysis, which used Monte Carlo simulations (see A5).
‡ Uncertainty accounted for using PPR from primary analysis and point death % from Zwaan[25] & Singh[26] combined.

C3. Triangulation of Available Data across Sources and Methods

We have described three separate methods of estimation that all yield compatible results:

- Method 1 (Manuscript): disease incidence x literature-based misdiagnosis-related harm rate
- Method 2 (Section C1): hospital deaths x % of deaths attributable to diagnostic error
- Method 3 (Section C2): hospital adverse events x % of adverse events resulting in harm or death

The three distinct methods can be used to derive inpatient serious harms and misdiagnosis-related deaths. Point estimates for total inpatient serious harms across the three methods range from ~135,000-225,000. All three methods produce point estimates for deaths that fall within the tight range of ~72,000-105,000.

Method 1 gives us total serious misdiagnosis-related harms (i.e., death + disability) in 2014 for inpatient and outpatient settings (~795,000). From this, we can estimate total deaths using the previously published proportion of harms representing deaths across inpatient and outpatient settings—46.7%[25,26] (~371,000) (see C1). Combining this with the proportion of total serious harms attributed to inpatient settings from a large, nationally representative sample of malpractice claims (28%[2]) gives ~225,000 total serious harms and ~105,000 deaths in US hospitals annually. Because of likely bias towards legal action for inpatient claims (see C2), these are presumed to be slight overestimates. Despite this, they are still close to results estimated by Methods 2 and 3, below. Method 1 serious harms (~225,000) fall within the uncertainty range by Method 2 (~96K-241K) and just beyond that by Method 3 (~97K-213K). Method 1 deaths (~105,000) fall within the uncertainty ranges by Method 2 (~51K-128K) and Method 3 (~51K-113K).

Method 2 gives us total misdiagnosis-related hospital deaths directly. Hospital deaths in 2014 (~980,000 [see C2]) were published by the CDC and the misdiagnosis-attributable fraction (8.4% [95% CI 5.2-13.1]) is from a rigorous meta-analysis of 53 autopsy studies whose final estimates account for study country, study epoch, and submaximal autopsy rate.[28] The estimate is ~82,000 (~51K-128K) misdiagnosis-related hospital inpatient deaths. Using the previously published proportion of serious harms representing deaths in the inpatient setting (~53.2% [29.1% deaths of ~54.7% serious harms][25]), we can estimate total serious harms of ~155,000 (~96K-241K). These are close to results from Method 3, despite different derivations.

Method 3 gives us total misdiagnosis-related hospital harms or deaths. US hospitalizations in 2014 (~35.4M[24]) derive from the NIS. The diagnostic adverse event rate (0.7%) is from a meta-analysis of 22 studies of hospital-based diagnostic adverse events,[33] and the proportion of adverse events resulting in serious harms (~54.7%) or death (29.1%) are from a rigorous, population-based sample of inpatient diagnostic adverse events from 21 hospitals. The resulting hospital estimate is ~135,000 (~97K-213K) serious misdiagnosis-related harms, which includes an estimated ~72,000 deaths (~51K-113K) (see C2).

Thus, our current results triangulate well across data sources and methods (convergent construct validity). When the consistency of these estimated misdiagnosis-related harms is combined with the consistent proportion of serious harms (~75-80%[3]) accounted for by "Big Three" diseases across settings (n=44 studies in primary care, emergency department, hospital[3]), this enhances the validity of our study results.

Finally, our results are bolstered by coherence with another recent systematic review and meta-analysis of diagnostic errors in the emergency department,[5] which permits a rough estimate across clinical settings as a final cross-check. As shown in Table S4, data combined from other sources (~855,000 [plausible range ~490K-1,659K]) align well with those found in the current study (~795,000 [PPR 598K-1,023K]). Both values translate to a US per-healthcare-visit serious misdiagnosis-related harm rate of about 0.08%.

Table S4. Serious misdiagnosis-related harms and serious harms rates by clinical setting (alternate sources)

Clinical Setting in which Error Occurs	Annual US Visits per Year in 2014 (n)	Total Serious Misdiagnosis-Related Harms (n)	Estimated Serious Harms Rate (%)
Inpatient	35,400,000[24]	~135,000 (~97K-213K)[25,33]*	~0.38% (~0.27-0.60)*
Emergency Department	137,800,000[48]	~430,000† (~259K-1,042K)†	~0.31% (~0.19-0.76)[5]
Primary Care Clinics	461,800,000[49]	~206,000† (~103K-309K)‡	~0.04%§ (~0.02-0.07)‡
Specialty Care Clinics	423,000,000[49]	~85,000† (~42K-127K)‡	~0.02%§ (~0.01-0.03)‡
TOTAL (£)	1,057,900,000	~855,000 (~490K-1,659K)	~0.08% (~0.05-0.16)

* Point estimate and uncertainty combine US inpatient harms estimate from Gunderson[33] with serious harms proportion from Zwaan[25] as described in Section C3 above. Harms rate is then calculated using visits per year.

† Total serious misdiagnosis-related harms calculated as serious harms rate x annual US visits.

‡ When precise estimates of uncertainty were lacking, we arbitrarily assigned it as +/- 50% of the point estimate.

§ Although the rate of serious harms from diagnostic error in ambulatory clinic-based care is not well characterized, generally the risks of a serious harm event (on a per visit basis) are much lower than either inpatient or emergency department care, simply because the severity of illness is much lower. From data in patients with missed stroke (erroneously called "benign" dizziness), we can estimate that the risks of serious harms after discharge from primary care are approximately 7-fold lower than those seen in the emergency department[50]; we can also approximate that the rate of serious harms after discharge from specialty care is likely about half that seen in primary care.[51]

£ Totals were calculated by summing n's for visits and harms, then dividing harms by visits to get rates (weighted average). As a result, the lower and upper uncertainty bounds are wider than if they were sampled probabilistically.

D) Supplementary References *(N.B. – citation numbers differ from main manuscript)*

1. Shojania KG, Dixon-Woods M. Estimating deaths due to medical error: the ongoing controversy and why it matters. *BMJ Qual Saf.* 2017;26(5):423-428.

2. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis (Berl).* 2019;6(3):227-240.

3. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Corrigendum to: Serious misdiagnosis-related harms in malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis (Berl).* 2020;8(1):127-128.

4. 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).* 2020;8(1):67-84.

5. Newman-Toker DE, Peterson SM, Badihian S, Hassoon A, Nassery N, Parizadeh D, Wilson LM, Jia Y, Omron R, Tharmarajah S, Guerin L, Bastani PB, Fracica EA, Kotwal S, Robinson KA. *Diagnostic Errors in the Emergency Department: A Systematic Review*. Comparative Effectiveness Review No. 258. (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No. 75Q80120D00003.) AHRQ Publication No. 22(23)-EHC043. Rockville, MD: Agency for Healthcare Research and Quality; December 2022 (addendum in press, anticipated August 2023). DOI: https://doi.org/10.23970/AHRQEPCCER258.

6. Localio AR, Lawthers AG, Brennan TA, et al. Relation between malpractice claims and adverse events due to negligence. Results of the Harvard Medical Practice Study III. *N Engl J Med.* 1991;325(4):245-251.

7. Studdert DM, Mello MM, Gawande AA, et al. Claims, errors, and compensation payments in medical malpractice litigation. *N Engl J Med.* 2006;354(19):2024-2033.

8. Producing National HCUP Estimates - Accessible Version. Agency for Healthcare Research and Quality, 2015. (Accessed May 20, 2023 at https://www.hcup-us.ahrq.gov/tech_assist/nationalestimates/508_course/508course_2018.jsp.)

9. National Trends in Inpatient Stays. Healthcare Cost and Utilization Project (HCUP) Fastats. Agency for Healthcare Research & Quality, 2023. (Accessed May 20, 2023 at https://datatools.ahrq.gov/hcup-fast-stats; https://datatools.ahrq.gov/sites/default/files/2022-10/DownloadTable_NationalTrendsInpatientStays_10-3-22.xls.)

10. Fast Facts on U.S. Hospitals, 2022. American Hospital Association. (Accessed May 20, 2023 at https://www.aha.org/statistics/fast-facts-us-hospitals.)

11. Meltzer DO, Hoomans T, Chung JW, Basu A. Minimal modeling approaches to value of information analysis for health research. *Med Decis Making.* 2011;31(6):E1-E22.

12. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* 1970;57(1):97–109.

13. Bartlett MS. The use of transformations. *Biometrics.* 1947;3(1):39-52.

14. ICD-10-CM Official Guidelines for Coding and Reporting FY 2019. Centers for Medicare & Medicaid Services, 2019. (Accessed May 20, 2023 at https://www.cms.gov/Medicare/Coding/ICD10/Downloads/2019-ICD10-Coding-Guidelines-.pdf.)

15. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One.* 2014;9(3):e92286.

16. Amann U, Heier M, Thilo C, Linseisen J, Meisinger C. Determinants of prehospital coronary heart disease death. *Sci Rep.* 2021;11(1):17134.

17. Interactive Atlas of Heart Disease and Stroke. Centers for Disease Control and Prevention. (Accessed May 20, 2023 at https://nccd.cdc.gov/DHDSPAtlas/reports.aspx?geographyType=county&state=NY&themeId=20&filterIds=1,5,3,6,7&filterOptions=6,1,1,1,1#report.)

18. Annual Report of Vital Statistics: New York State 2014. New York State Department of Health. (Accessed May 20, 2023 at https://www.health.ny.gov/statistics/vital_statistics/docs/vital_statistics_annual_report_2014.pdf.)

19. Stroke Facts. Centers for Disease Control and Prevention. (Accessed May 20, 2023 at https://www.cdc.gov/stroke/facts.htm.)

20. Wales K. Pre-Transport Stroke Death Rates by EMS Region and Location of Stroke Designated Hospitals, New York State. Centers for Disease Control and Prevention. (Accessed May 20, 2023 at https://www.cdc.gov/dhdsp/maps/gisx/mapgallery/NY-stroke.html.)

21. O'Neill L. Estimating out-of-hospital mortality due to myocardial infarction. *Health Care Manag Sci.* 2003;6(3):147-154.

22. Giles MF, Flossman E, Rothwell PM. Patient behavior immediately after transient ischemic attack according to clinical characteristics, perception of the event, and predicted risk of stroke. *Stroke.* 2006;37(5):1254-1260.

23. Summary Health Statistics: National Health Interview Survey, 2014. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2014_SHS_Table_P-10.pdf.)

24. Trends in Hospital Inpatient Stays in the United States, 2005-2014 (HCUP Statistical Brief #225). Agency for Healthcare Research & Quality. (Accessed May 20, 2023 at https://www.hcup-us.ahrq.gov/reports/statbriefs/sb225-Inpatient-US-Stays-Trends.jsp.)

25. Zwaan L, de Bruijne M, Wagner C, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med.* 2010;170(12):1015-1021.

26. 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(6):418-425.

27. Deaths: Final Data for 2014. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://www.cdc.gov/nchs/data/dvs/LCWK1_2014.pdf.)

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

29. Kowalski RG, Claassen J, Kreiter KT, et al. Initial misdiagnosis and outcome after subarachnoid hemorrhage. *JAMA.* 2004;291(7):866-869.

30. 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(15):1468-1477.

31. Singh H, Giardina TD, Forjuoh SN, et al. Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ Qual Saf.* 2012;21(2):93-100.

32. 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(9):727-731.

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

34. Ambulatory Care Use and Physician Office Visits. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at http://www.cdc.gov/nchs/fastats/physician-visits.htm.)

35. Emergency Department Visits. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at http://www.cdc.gov/nchs/fastats/emergency-department.htm.)

36. Trends in Inpatient Hospital Deaths: National Hospital Discharge Survey, 2000–2010. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://www.cdc.gov/nchs/products/databriefs/db118.htm.)

37. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA.* 2001;286(4):415-420.

38. From a Small Scrape to Sepsis - Rory Staunton's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/from-a-small-scrape-to-sepsis/.)

39. Unbelieved - Jessica Barnett's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/unbelieved-jessica-barnetts-story/.)

40. Tested Too Late - Chad Becken's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/tested-too-late-chad-beckens-story/.)

41. Missed Stroke Diagnosis - John Michael Night's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/missed-stroke-diagnosis/.)

42. The Missed Bilirubin Test - Cal Sheridan's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/the-missed-bilirubin-test-cal-sheridans-story/.)

43. The Lingering Lump - Diana Cejas' Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/the-lingering-lump-diana-cejas-story/.)

44. Smidfelt K, Nordanstig J, Davidsson A, Torngren K, Langenskiold M. Misdiagnosis of ruptured abdominal aortic aneurysms is common and is associated with increased mortality. *J Vasc Surg.* 2021;73(2):476-483 e473.

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

46. Leape LL, Berwick DM, Bates DW. Counting deaths due to medical errors (in reply). *JAMA.* 2002;288(19):2404-2405.

47. QuickStats: Percentage Distribution of Deaths, by Place of Death — United States, 2000–2014. *Morbidity and Mortality Weekly Report.* 2016;65:357. (Accessed May 20, 2023 at https://www.cdc.gov/mmwr/volumes/65/wr/mm6513a6.htm.)

48. Hooker EA, Mallow PJ, Oglesby MM. Characteristics and Trends of Emergency Department Visits in the United States (2010-2014). *J Emerg Med.* 2019;56(3):344-351.

49. National Ambulatory Medical Care Survey: 2014 State and National Summary Tables. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2014_namcs_web_tables.pdf.)

50. Nassery N, Mane K, Liu F, et al. Measuring missed strokes using administrative and claims data: towards a diagnostic performance dashboard to monitor diagnostic errors. Diagnostic Error in Medicine 2016; November 6-8, 2016, 2016; Hollywood, CA.

51. Chang TP, Bery AK, Wang Z, et al. Stroke hospitalization after misdiagnosis of "benign dizziness" is lower in specialty care than general practice: a population-based cohort analysis of missed stroke using SPADE methods. *Diagnosis (Berl).* 2021;9(1):96-106.

Supplemental File #2 for ***Burden of Serious Harms from Diagnostic Error in the USA***
by Newman-Toker, et al. BMJ Quality & Safety, 2023 (doi:10.1136/bmjqs-2021-014130)

Enclosed below are additional Supplementary Materials (Supplement 2) to the peer-reviewed, scientific journal publication entitled ***Burden of Serious Harms from Diagnostic Error in the USA*** *(2023).*

This is the full statistical code (R v4.2.2, Vienna, Austria) for the Monte Carlo analysis.

```
# load packages
rm(list = ls())
library(sn)
library(rootSolve)
library(gdata)
library(MASS)
library(DEoptim)

set.seed(37)
logit <- function(x) {
    out <- log(x/(1-x))
    return(out)
}

expit <- function(x) {
    out <- exp(x) / (1 + exp(x))
    return(out)
}

#### primary analysis  ####
dat <- read.csv(file = "big3_data_dx1.csv")

peh_general <- 374/1216 # general per-error-harm rate
```

```
peh_general_se <- sqrt(1/374 + 1/(1216-374)) # standard error on logit scale
peh_general_lower <- expit(logit(peh_general) + qnorm(0.025) * peh_general_se)
peh_general_upper <- expit(logit(peh_general) + qnorm(0.975) * peh_general_se)

perror_big3 <- 717 / (717 + 509) # proportion of big 3 in errors
pharm_big3 <- 157 / 207 # proportion of big 3 in harms

vascular_index <- which(dat$category == "vascular" & dat$disease != "OTHER Vascular")
vascular_other_index <- which(dat$disease == "OTHER Vascular")
infection_index <- which(dat$category == "infection" & dat$disease != "OTHER Infection")
infection_other_index <- which(dat$disease == "OTHER Infection")
cancer_index <- which(dat$category == "cancer" & dat$disease != "OTHER Cancer")
cancer_other_index <- which(dat$disease == "OTHER Cancer")

dat$nerror <- dat$incidence_point * dat$misrate_point
shp <- dat$proportion_highharm # severity harm proportion

peh_fun <- function(peh_general, nerror) {
 # function to calculate disease-specific per-error-harm rates
   peh_rate <- rep(NA, nrow(dat))
   peh_rate[c(vascular_index, infection_index, cancer_index)] <- peh_general * pharm_big3 /
perror_big3 *
       (sum(nerror) * shp[c(vascular_index, infection_index, cancer_index)]) /
       (sum(nerror[vascular_index] * shp[vascular_index]) * 1684 / 1344 +
           sum(nerror[infection_index] * shp[infection_index]) * 992 / 600 +
           sum(nerror[cancer_index] * shp[cancer_index]) * 2793 / 1529)
   peh_rate[vascular_other_index] <- sum(nerror[vascular_index] * peh_rate[vascular_index] *
334/1334) /
       nerror[vascular_other_index]
```

```
    peh_rate[infection_other_index] <- sum(nerror[infection_index] * peh_rate[infection_index] *
392/600) /
        nerror[infection_other_index]
    peh_rate[cancer_other_index] <- sum(nerror[cancer_index] * peh_rate[cancer_index] * 1264/1529)
/
        nerror[cancer_other_index]
    return(peh_rate)
}

# point estimate of per-error-serious harm rates
dat$peharmrate_point <- peh_fun(peh_general, dat$nerror)
dat$peharmrate_lb <- peh_fun(peh_general_lower, dat$nerror)
dat$peharmrate_ub <- peh_fun(peh_general_upper, dat$nerror)

# Monte Carlo simulation, to construct plausible intervals for harm rate and number of harm
snest <- function(u, lower, upper, lowerq = 0.025, upperq = 0.975) {
    solfun <- function(beta) {
        omega <- beta[1]
        alpha <- beta[2]
        out1 <- psn(lower, xi = u, omega = omega, alpha = alpha) - lowerq
        out2 <- psn(upper, xi = u, omega = omega, alpha = alpha) - upperq
        return((out1^2+out2^2))
    }
    start <- c(((upper-lower)/4)^2, 0) + mvrnorm(100, mu = rep(0, 2), Sigma = diag(1, 2))
    start[, 1] <- abs(start[, 1])
    tempfit <- t(apply(start, 1, function(x) {
        fit <- optim(fn = solfun, par = x,
                control = list(maxit = 500000))
        return(c(fit$par, fit$value))
```

```
    }))
    tempfit <- tempfit[complete.cases(tempfit), ]
    return(tempfit[which(tempfit[, 3] == min(tempfit[, 3]))[1], 1:3])
}



totl <- 316283434
nmc <- 1*(10^7)
numerical_error <- NULL
dat$nerror_lb <- NA
dat$nerror_ub <- NA
dat$nharm_point <- dat$nerror * dat$peharmrate_point
dat$nharm_lb <- NA
dat$nharm_ub <- NA
dat$harmrate_point <- dat$misrate_point * dat$peharmrate_point
dat$harmrate_lb <- NA
dat$harmrate_ub <- NA

# Monte Carlo simulation based on skew-normal estimation
mc_result <- list()
for (i in 1:nrow(dat)) {
    mc_result[[i]] <- list()
    if (dat$incidence_lb[i] == dat$incidence_ub[i]) {
        mc_incidence <- rep(dat$incidence_point[i], nmc)
    } else {
        fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                               lower = logit(dat$incidence_lb[i] / totl),
                               upper = logit(dat$incidence_ub[i] / totl))
        while(fit_incidence[3] > 10^(-10)) {
```

```
            fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                                   lower = logit(dat$incidence_lb[i] / totl),
                                   upper = logit(dat$incidence_ub[i] / totl))
        }
        mc_incidence <- expit(rsn(n = nmc, xi = logit(dat$incidence_point[i] / totl),
                                  omega = fit_incidence[1], alpha = fit_incidence[2])) * totl
    }
    temp <- (c(quantile(mc_incidence, c(0.025, 0.975)) - c(dat$incidence_lb[i], dat$incidence_ub[i]))) /
totl

    if (i %in% c(vascular_index, infection_index, cancer_index)) {
        fit_misrate <- snest(logit(dat$misrate_point[i]),
                             lower = logit(dat$misrate_lb[i]),
                             upper = logit(dat$misrate_ub[i]))
        while(fit_misrate[3] > 10^(-10)) {
            fit_misrate <- snest(logit(dat$misrate_point[i]),
                                 lower = logit(dat$misrate_lb[i]),
                                 upper = logit(dat$misrate_ub[i]))
        }

        mc_misrate <- expit(rsn(n = nmc, xi = logit(dat$misrate_point[i]),
                                omega = fit_misrate[1], alpha = fit_misrate[2]))
        temp <- c(temp, quantile(mc_misrate, c(0.025, 0.975)) - c(dat$misrate_lb[i], dat$misrate_ub[i]))

        mc_result[[i]]$fit_misrate <- fit_misrate
        mc_result[[i]]$mc_misrate <- mc_misrate
    } else {
        temp <- c(temp, rep(NA, 2))
    }
```

```
    fit_peh <- snest(logit(dat$peharmrate_point[i]),
                     lower = logit(dat$peharmrate_lb[i]),
                     upper = logit(dat$peharmrate_ub[i]))
    while(fit_peh[3] > 10^(-10)) {
        fit_peh <- snest(logit(dat$peharmrate_point[i]),
                         lower = logit(dat$peharmrate_lb[i]),
                         upper = logit(dat$peharmrate_ub[i]))
    }
    mc_peh <- expit(rsn(n = nmc, xi = logit(dat$peharmrate_point[i]),
                        omega = fit_peh[1], alpha = fit_peh[2]))
    temp <- c(temp, quantile(mc_peh, c(0.025, 0.975)) - c(dat$peharmrate_lb[i], dat$peharmrate_ub[i]))

    numerical_error <- rbind(numerical_error, temp)

    mc_result[[i]]$fit_incidence <- fit_incidence
    mc_result[[i]]$mc_incidence <- mc_incidence

    mc_result[[i]]$fit_peh <- fit_peh
    mc_result[[i]]$mc_peh <- mc_peh
    mc_result[[i]]$numerical_error <- numerical_error
}

summary(numerical_error)

# calculate plausible intervals for "OTHER" vascular/infection/censor harm rate (misrate)
# because point estimate is calculated as weighted sum and plausible intervals should be calcualted accordingly
temp1 <- 0
```

```
temp2 <- 0
for (i in vascular_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[vascular_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate,
0.025)
dat$misrate_ub[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate,
0.975)



temp1 <- 0
temp2 <- 0
for (i in infection_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[infection_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate,
0.025)
dat$misrate_ub[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate,
0.975)


temp1 <- 0
temp2 <- 0
for (i in cancer_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
```

```
mc_result[[cancer_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.975)



# calculate plausible intervals for number of errors, number of harms and harm rate
# based on Monte Carlo generated data
for (i in 1:nrow(dat)) {
    mc_incidence <- mc_result[[i]]$mc_incidence
    mc_misrate <- mc_result[[i]]$mc_misrate
    mc_peh <- mc_result[[i]]$mc_peh
    dat$nerror_lb[i] <- quantile(mc_incidence * mc_misrate, 0.025)
    dat$nerror_ub[i] <- quantile(mc_incidence * mc_misrate, 0.975)

    dat$nharm_lb[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.025)
    dat$nharm_ub[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.975)

    dat$harmrate_lb[i] <- quantile(mc_misrate * mc_peh, 0.025)
    dat$harmrate_ub[i] <- quantile(mc_misrate * mc_peh, 0.975)
}

# calculate subtotals and plausible intervals
# create structure for subtotals
temp <- c("Top 5 Vascular Subtotal",
        "Top 5 Infection Subtotal",
        "Top 5 Cancer Subtotal",
        "Top 15 Subtotal",
        "Total Vascular",
```

```
        "Total Infection",
        "Total Cancer",
        "Big 3 Subtotal",
        "All Other Dx Errors",
        "Grand Total")
temp2 <- data.frame(matrix(NA, nrow = length(temp), ncol = ncol(dat)))
colnames(temp2) <- colnames(dat)
temp2[, 2] <- temp
temp2[, 1] <- rep("Subtotals", length(temp))
dat <- rbind(dat, temp2)

dat[, -c(1:2)] <- apply(dat[, -c(1:2)], c(1, 2), as.numeric)

# calculate subtotal incidence
dat$incidence_point[19] <- sum(dat$incidence_point[1:5])
dat$incidence_point[20] <- sum(dat$incidence_point[7:11])
dat$incidence_point[21] <- sum(dat$incidence_point[13:17])
dat$incidence_point[22] <- sum(dat$incidence_point[19:21])

dat$incidence_point[23] <- sum(dat$incidence_point[1:6])
dat$incidence_point[24] <- sum(dat$incidence_point[7:12])
dat$incidence_point[25] <- sum(dat$incidence_point[13:18])
dat$incidence_point[26] <- sum(dat$incidence_point[1:18])


temp <-lapply(1:18, function(i) return(mc_result[[i]]$mc_incidence))

temp1 <- rep(0, nmc)
for (i in 1:5) {
```

```
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[19] <- quantile(temp1, 0.025)
dat$incidence_ub[19] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[20] <- quantile(temp1, 0.025)
dat$incidence_ub[20] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[21] <- quantile(temp1, 0.025)
dat$incidence_ub[21] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[22] <- quantile(temp1, 0.025)
dat$incidence_ub[22] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
```

```
for (i in 1:6) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[23] <- quantile(temp1, 0.025)
dat$incidence_ub[23] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[24] <- quantile(temp1, 0.025)
dat$incidence_ub[24] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[25] <- quantile(temp1, 0.025)
dat$incidence_ub[25] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[26] <- quantile(temp1, 0.025)
dat$incidence_ub[26] <- quantile(temp1, 0.975)
```

```
dat[27:28, 3:5] <- NA

# calculate subtotal number of errors
dat$nerror[19] <- sum(dat$nerror[1:5])
dat$nerror[20] <- sum(dat$nerror[7:11])
dat$nerror[21] <- sum(dat$nerror[13:17])
dat$nerror[22] <- sum(dat$nerror[19:21])

dat$nerror[23] <- sum(dat$nerror[1:6])
dat$nerror[24] <- sum(dat$nerror[7:12])
dat$nerror[25] <- sum(dat$nerror[13:18])
dat$nerror[26] <- sum(dat$nerror[1:18])

dat$nerror[27] <- dat$nerror[26] / perror_big3 * (1 - perror_big3)
dat$nerror[28] <- dat$nerror[26] / perror_big3



temp1 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[19] <- quantile(temp1, 0.025)
dat$nerror_ub[19] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
```

```
}
dat$nerror_lb[20] <- quantile(temp1, 0.025)
dat$nerror_ub[20] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[21] <- quantile(temp1, 0.025)
dat$nerror_ub[21] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[22] <- quantile(temp1, 0.025)
dat$nerror_ub[22] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[23] <- quantile(temp1, 0.025)
dat$nerror_ub[23] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
```

```
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[24] <- quantile(temp1, 0.025)
dat$nerror_ub[24] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[25] <- quantile(temp1, 0.025)
dat$nerror_ub[25] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[26] <- quantile(temp1, 0.025)
dat$nerror_ub[26] <- quantile(temp1, 0.975)

dat$nerror_lb[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.025)
dat$nerror_ub[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.975)

dat$nerror_lb[28] <- quantile(temp1 / perror_big3, 0.025)
dat$nerror_ub[28] <- quantile(temp1 / perror_big3, 0.975)
```

```
# calculate subtotal misrate
dat$misrate_point[19:28] <- dat$nerror[19:28] / dat$incidence_point[19:28]

temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[19] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[19] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[20] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[20] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
```

```
dat$misrate_lb[21] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[21] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[22] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[22] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[23] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[23] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
```

```
}
dat$misrate_lb[24] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[24] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[25] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[25] <- quantile(temp1 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[26] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[26] <- quantile(temp1 / temp2, 0.975)



# calculate subtotal nharm, peh, and harm rate
dat$nharm_point[19] <- sum(dat$nharm_point[1:5])
dat$nharm_point[20] <- sum(dat$nharm_point[7:11])
```

```
dat$nharm_point[21] <- sum(dat$nharm_point[13:17])
dat$nharm_point[22] <- sum(dat$nharm_point[19:21])

dat$nharm_point[23] <- sum(dat$nharm_point[1:6])
dat$nharm_point[24] <- sum(dat$nharm_point[7:12])
dat$nharm_point[25] <- sum(dat$nharm_point[13:18])
dat$nharm_point[26] <- sum(dat$nharm_point[1:18])

dat$nharm_point[27] <- dat$nharm_point[26] / pharm_big3 * (1 - pharm_big3)
dat$nharm_point[28] <- dat$nharm_point[26] / pharm_big3



dat$harmrate_point[19:28] <- dat$nharm_point[19:28] / dat$incidence_point[19:28]
dat$peharmrate_point[19:28] <- dat$nharm_point[19:28] / dat$nerror[19:28]



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[19] <- quantile(temp3, 0.025)
dat$nharm_ub[19] <- quantile(temp3, 0.975)

dat$harmrate_lb[19] <- quantile(temp3 / temp1, 0.025)
```

```
dat$harmrate_ub[19] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[19] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[19] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[20] <- quantile(temp3, 0.025)
dat$nharm_ub[20] <- quantile(temp3, 0.975)

dat$harmrate_lb[20] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[20] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[20] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[20] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[21] <- quantile(temp3, 0.025)
dat$nharm_ub[21] <- quantile(temp3, 0.975)

dat$harmrate_lb[21] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[21] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[21] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[21] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
```

```
}

dat$nharm_lb[22] <- quantile(temp3, 0.025)
dat$nharm_ub[22] <- quantile(temp3, 0.975)

dat$harmrate_lb[22] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[22] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[22] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[22] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[23] <- quantile(temp3, 0.025)
dat$nharm_ub[23] <- quantile(temp3, 0.975)

dat$harmrate_lb[23] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[23] <- quantile(temp3 / temp1, 0.975)
```

```
dat$peharmrate_lb[23] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[23] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[24] <- quantile(temp3, 0.025)
dat$nharm_ub[24] <- quantile(temp3, 0.975)

dat$harmrate_lb[24] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[24] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[24] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[24] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
```

```
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[25] <- quantile(temp3, 0.025)
dat$nharm_ub[25] <- quantile(temp3, 0.975)

dat$harmrate_lb[25] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[25] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[25] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[25] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}
```

```
dat$nharm_lb[26] <- quantile(temp3, 0.025)
dat$nharm_ub[26] <- quantile(temp3, 0.975)

dat$harmrate_lb[26] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[26] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[26] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[26] <- quantile(temp3 / temp2, 0.975)



dat$nharm_lb[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.025)
dat$nharm_ub[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.975)

dat$nharm_lb[28] <- quantile(temp3 / pharm_big3, 0.025)
dat$nharm_ub[28] <- quantile(temp3 / pharm_big3, 0.975)



dat$peharmrate_lb[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                                  (temp2 / perror_big3 * (1 - perror_big3)), 0.025)
dat$peharmrate_ub[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                                  (temp2 / perror_big3 * (1 - perror_big3)), 0.975)

dat$peharmrate_lb[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.025)
dat$peharmrate_ub[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.975)

dat$weight <- dat$peharmrate_point / peh_general
dat <- dat[, c(1:2, 9:10, 23, 3:5,
        6:8, 12:14, 20:22,
        11, 15, 16, 17:19)]
```

```
all(dat$misrate_lb < dat$misrate_point, na.rm = T)
all(dat$misrate_ub > dat$misrate_point, na.rm = T)
all(dat$harmrate_lb < dat$harmrate_point, na.rm = T)
all(dat$harmrate_ub > dat$harmrate_point, na.rm = T)
all(dat$nharm_lb < dat$nharm_point, na.rm = T)
all(dat$nharm_ub > dat$nharm_point, na.rm = T)
all(dat$nerror_lb < dat$nerror, na.rm = T)
all(dat$nerror_ub > dat$nerror, na.rm = T)

write.csv(dat, file = "big3_results_dx1_updated.csv")



#### secondary analysis  ####
dat <- read.csv(file = "big3_data_combo.csv")
set.seed(37)
peh_general <- 374/1216 # general per-error-harm rate
peh_general_se <- sqrt(1/374 + 1/(1216-374)) # standard error on logit scale
peh_general_lower <- expit(logit(peh_general) + qnorm(0.025) * peh_general_se)
peh_general_upper <- expit(logit(peh_general) + qnorm(0.975) * peh_general_se)

perror_big3 <- 717 / (717 + 509) # big 3 error rate
pharm_big3 <- 157 / 207 # harm rate

vascular_index <- which(dat$category == "vascular" & dat$disease != "OTHER Vascular")
vascular_other_index <- which(dat$disease == "OTHER Vascular")
infection_index <- which(dat$category == "infection" & dat$disease != "OTHER Infection")
```

```
infection_other_index <- which(dat$disease == "OTHER Infection")

cancer_index <- which(dat$category == "cancer" & dat$disease != "OTHER Cancer")

cancer_other_index <- which(dat$disease == "OTHER Cancer")


dat$nerror <- dat$incidence_point * dat$misrate_point

shp <- dat$proportion_highharm # severity harm proportion


peh_fun <- function(peh_general, nerror) {

 # function to calculate disease-specific per-error-harm rates

    peh_rate <- rep(NA, nrow(dat))

    peh_rate[c(vascular_index, infection_index, cancer_index)] <- peh_general * pharm_big3 /
perror_big3 *

        (sum(nerror) * shp[c(vascular_index, infection_index, cancer_index)]) /

        (sum(nerror[vascular_index] * shp[vascular_index]) * 1684 / 1344 +

            sum(nerror[infection_index] * shp[infection_index]) * 992 / 600 +

            sum(nerror[cancer_index] * shp[cancer_index]) * 2793 / 1529)

    peh_rate[vascular_other_index] <- sum(nerror[vascular_index] * peh_rate[vascular_index] *
334/1334) /

        nerror[vascular_other_index]

    peh_rate[infection_other_index] <- sum(nerror[infection_index] * peh_rate[infection_index] *
392/600) /

        nerror[infection_other_index]

    peh_rate[cancer_other_index] <- sum(nerror[cancer_index] * peh_rate[cancer_index] * 1264/1529)
/

        nerror[cancer_other_index]

    return(peh_rate)

}


# point estimate of per-error-serious harm rates

dat$peharmrate_point <- peh_fun(peh_general, dat$nerror)
```

```
dat$peharmrate_lb <- peh_fun(peh_general_lower, dat$nerror)
dat$peharmrate_ub <- peh_fun(peh_general_upper, dat$nerror)

# Monte Carlo simulation, to construct plausible intervals for harm rate and number of harm
snest <- function(u, lower, upper, lowerq = 0.025, upperq = 0.975) {
    solfun <- function(beta) {
        omega <- beta[1]
        alpha <- beta[2]
        out1 <- psn(lower, xi = u, omega = omega, alpha = alpha) - lowerq
        out2 <- psn(upper, xi = u, omega = omega, alpha = alpha) - upperq
        return((out1^2+out2^2))
    }
    start <- c(((upper-lower)/4)^2, 0) + mvrnorm(100, mu = rep(0, 2), Sigma = diag(1, 2))
    start[, 1] <- abs(start[, 1])
    tempfit <- t(apply(start, 1, function(x) {
        fit <- optim(fn = solfun, par = x,
                control = list(maxit = 500000))
        return(c(fit$par, fit$value))
    }))
    tempfit <- tempfit[complete.cases(tempfit), ]
    return(tempfit[which(tempfit[, 3] == min(tempfit[, 3]))[1], 1:3])
}

totl <- 316283434
nmc <- 1*(10^7)
numerical_error <- NULL
dat$nerror_lb <- NA
dat$nerror_ub <- NA
dat$nharm_point <- dat$nerror * dat$peharmrate_point
```

```
dat$nharm_lb <- NA
dat$nharm_ub <- NA
dat$harmrate_point <- dat$misrate_point * dat$peharmrate_point
dat$harmrate_lb <- NA
dat$harmrate_ub <- NA

# Monte Carlo simulation based on skew-normal estimation
mc_result <- list()
for (i in 1:nrow(dat)) {
    mc_result[[i]] <- list()
    if (dat$incidence_lb[i] == dat$incidence_ub[i]) {
        mc_incidence <- rep(dat$incidence_point[i], nmc)
    } else {
        fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                               lower = logit(dat$incidence_lb[i] / totl),
                               upper = logit(dat$incidence_ub[i] / totl))
        while(fit_incidence[3] > 10^(-10)) {
            fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                                   lower = logit(dat$incidence_lb[i] / totl),
                                   upper = logit(dat$incidence_ub[i] / totl))
        }
        mc_incidence <- expit(rsn(n = nmc, xi = logit(dat$incidence_point[i] / totl),
                                  omega = fit_incidence[1], alpha = fit_incidence[2])) * totl
    }
    temp <- (c(quantile(mc_incidence, c(0.025, 0.975)) - c(dat$incidence_lb[i], dat$incidence_ub[i]))) /
totl

    if (i %in% c(vascular_index, infection_index, cancer_index)) {
        fit_misrate <- snest(logit(dat$misrate_point[i]),
```

```
                    lower = logit(dat$misrate_lb[i]),
                    upper = logit(dat$misrate_ub[i]))
    while(fit_misrate[3] > 10^(-10)) {
        fit_misrate <- snest(logit(dat$misrate_point[i]),
                    lower = logit(dat$misrate_lb[i]),
                    upper = logit(dat$misrate_ub[i]))
    }

    mc_misrate <- expit(rsn(n = nmc, xi = logit(dat$misrate_point[i]),
                    omega = fit_misrate[1], alpha = fit_misrate[2]))
    temp <- c(temp, quantile(mc_misrate, c(0.025, 0.975)) - c(dat$misrate_lb[i], dat$misrate_ub[i]))

    mc_result[[i]]$fit_misrate <- fit_misrate
    mc_result[[i]]$mc_misrate <- mc_misrate
} else {
    temp <- c(temp, rep(NA, 2))
}

fit_peh <- snest(logit(dat$peharmrate_point[i]),
            lower = logit(dat$peharmrate_lb[i]),
            upper = logit(dat$peharmrate_ub[i]))
while(fit_peh[3] > 10^(-10)) {
    fit_peh <- snest(logit(dat$peharmrate_point[i]),
                lower = logit(dat$peharmrate_lb[i]),
                upper = logit(dat$peharmrate_ub[i]))
}
mc_peh <- expit(rsn(n = nmc, xi = logit(dat$peharmrate_point[i]),
                omega = fit_peh[1], alpha = fit_peh[2]))
temp <- c(temp, quantile(mc_peh, c(0.025, 0.975)) - c(dat$peharmrate_lb[i], dat$peharmrate_ub[i]))
```

```
    numerical_error <- rbind(numerical_error, temp)

    mc_result[[i]]$fit_incidence <- fit_incidence
    mc_result[[i]]$mc_incidence <- mc_incidence

    mc_result[[i]]$fit_peh <- fit_peh
    mc_result[[i]]$mc_peh <- mc_peh
    mc_result[[i]]$numerical_error <- numerical_error
}
summary(numerical_error) # 10e-5

# calculate plausible intervals for "OTHER" vascular/infection/censor harm rate (misrate)
# because point estimate is calculated as weighted sum and plausible intervals should be calcualted accordingly
temp1 <- 0
temp2 <- 0
for (i in vascular_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[vascular_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate, 0.975)


temp1 <- 0
temp2 <- 0
```

```
for (i in infection_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[infection_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate, 0.975)

temp1 <- 0
temp2 <- 0
for (i in cancer_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[cancer_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.975)

# calculate plausible intervals for number of errors, number of harms and harm rate
# based on Monte Carlo generated data
for (i in 1:nrow(dat)) {
    mc_incidence <- mc_result[[i]]$mc_incidence
    mc_misrate <- mc_result[[i]]$mc_misrate
    mc_peh <- mc_result[[i]]$mc_peh
    dat$nerror_lb[i] <- quantile(mc_incidence * mc_misrate, 0.025)
```

```
    dat$nerror_ub[i] <- quantile(mc_incidence * mc_misrate, 0.975)

    dat$nharm_lb[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.025)
    dat$nharm_ub[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.975)

    dat$harmrate_lb[i] <- quantile(mc_misrate * mc_peh, 0.025)
    dat$harmrate_ub[i] <- quantile(mc_misrate * mc_peh, 0.975)
}

# calculate subtotals and plausible intervals
# create structure for subtotals
temp <- c("Top 5 Vascular Subtotal",
        "Top 5 Infection Subtotal",
        "Top 5 Cancer Subtotal",
        "Top 15 Subtotal",
        "Total Vascular",
        "Total Infection",
        "Total Cancer",
        "Big 3 Subtotal",
        "All Other Dx Errors",
        "Grand Total")
temp2 <- data.frame(matrix(NA, nrow = length(temp), ncol = ncol(dat)))
colnames(temp2) <- colnames(dat)
temp2[, 2] <- temp
temp2[, 1] <- rep("Subtotals", length(temp))
dat <- rbind(dat, temp2)

dat[, -c(1:2)] <- apply(dat[, -c(1:2)], c(1, 2), as.numeric)
```

```
# calculate subtotal incidence
dat$incidence_point[19] <- sum(dat$incidence_point[1:5])
dat$incidence_point[20] <- sum(dat$incidence_point[7:11])
dat$incidence_point[21] <- sum(dat$incidence_point[13:17])
dat$incidence_point[22] <- sum(dat$incidence_point[19:21])

dat$incidence_point[23] <- sum(dat$incidence_point[1:6])
dat$incidence_point[24] <- sum(dat$incidence_point[7:12])
dat$incidence_point[25] <- sum(dat$incidence_point[13:18])
dat$incidence_point[26] <- sum(dat$incidence_point[1:18])



temp <-lapply(1:18, function(i) return(mc_result[[i]]$mc_incidence))

temp1 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[19] <- quantile(temp1, 0.025)
dat$incidence_ub[19] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[20] <- quantile(temp1, 0.025)
dat$incidence_ub[20] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
```

```
for (i in 13:17) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[21] <- quantile(temp1, 0.025)
dat$incidence_ub[21] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[22] <- quantile(temp1, 0.025)
dat$incidence_ub[22] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[23] <- quantile(temp1, 0.025)
dat$incidence_ub[23] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[24] <- quantile(temp1, 0.025)
dat$incidence_ub[24] <- quantile(temp1, 0.975)
```

```
temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[25] <- quantile(temp1, 0.025)
dat$incidence_ub[25] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[26] <- quantile(temp1, 0.025)
dat$incidence_ub[26] <- quantile(temp1, 0.975)

dat[27:28, 3:5] <- NA

# calculate subtotal number of errors
dat$nerror[19] <- sum(dat$nerror[1:5])
dat$nerror[20] <- sum(dat$nerror[7:11])
dat$nerror[21] <- sum(dat$nerror[13:17])
dat$nerror[22] <- sum(dat$nerror[19:21])

dat$nerror[23] <- sum(dat$nerror[1:6])
dat$nerror[24] <- sum(dat$nerror[7:12])
dat$nerror[25] <- sum(dat$nerror[13:18])
dat$nerror[26] <- sum(dat$nerror[1:18])
```

```
dat$nerror[27] <- dat$nerror[26] / perror_big3 * (1 - perror_big3)
dat$nerror[28] <- dat$nerror[26] / perror_big3



temp1 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[19] <- quantile(temp1, 0.025)
dat$nerror_ub[19] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[20] <- quantile(temp1, 0.025)
dat$nerror_ub[20] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[21] <- quantile(temp1, 0.025)
dat$nerror_ub[21] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
```

```
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[22] <- quantile(temp1, 0.025)
dat$nerror_ub[22] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[23] <- quantile(temp1, 0.025)
dat$nerror_ub[23] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[24] <- quantile(temp1, 0.025)
dat$nerror_ub[24] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[25] <- quantile(temp1, 0.025)
dat$nerror_ub[25] <- quantile(temp1, 0.975)
```

```
temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[26] <- quantile(temp1, 0.025)
dat$nerror_ub[26] <- quantile(temp1, 0.975)

dat$nerror_lb[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.025)
dat$nerror_ub[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.975)

dat$nerror_lb[28] <- quantile(temp1 / perror_big3, 0.025)
dat$nerror_ub[28] <- quantile(temp1 / perror_big3, 0.975)




# calculate subtotal misrate
dat$misrate_point[19:28] <- dat$nerror[19:28] / dat$incidence_point[19:28]

temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[19] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[19] <- quantile(temp1 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[20] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[20] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[21] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[21] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[22] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[22] <- quantile(temp1 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[23] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[23] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[24] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[24] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[25] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[25] <- quantile(temp1 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[26] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[26] <- quantile(temp1 / temp2, 0.975)


# calculate subtotal nharm, peh, and harm rate
dat$nharm_point[19] <- sum(dat$nharm_point[1:5])
dat$nharm_point[20] <- sum(dat$nharm_point[7:11])
dat$nharm_point[21] <- sum(dat$nharm_point[13:17])
dat$nharm_point[22] <- sum(dat$nharm_point[19:21])

dat$nharm_point[23] <- sum(dat$nharm_point[1:6])
dat$nharm_point[24] <- sum(dat$nharm_point[7:12])
dat$nharm_point[25] <- sum(dat$nharm_point[13:18])
dat$nharm_point[26] <- sum(dat$nharm_point[1:18])

dat$nharm_point[27] <- dat$nharm_point[26] / pharm_big3 * (1 - pharm_big3)
dat$nharm_point[28] <- dat$nharm_point[26] / pharm_big3


dat$harmrate_point[19:28] <- dat$nharm_point[19:28] / dat$incidence_point[19:28]
```

```
dat$peharmrate_point[19:28] <- dat$nharm_point[19:28] / dat$nerror[19:28]



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[19] <- quantile(temp3, 0.025)
dat$nharm_ub[19] <- quantile(temp3, 0.975)

dat$harmrate_lb[19] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[19] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[19] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[19] <- quantile(temp3 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
```

```
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[20] <- quantile(temp3, 0.025)
dat$nharm_ub[20] <- quantile(temp3, 0.975)

dat$harmrate_lb[20] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[20] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[20] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[20] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[21] <- quantile(temp3, 0.025)
dat$nharm_ub[21] <- quantile(temp3, 0.975)

dat$harmrate_lb[21] <- quantile(temp3 / temp1, 0.025)
```

```
dat$harmrate_ub[21] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[21] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[21] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[22] <- quantile(temp3, 0.025)
dat$nharm_ub[22] <- quantile(temp3, 0.975)

dat$harmrate_lb[22] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[22] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[22] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[22] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[23] <- quantile(temp3, 0.025)
dat$nharm_ub[23] <- quantile(temp3, 0.975)

dat$harmrate_lb[23] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[23] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[23] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[23] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
```

```
}

dat$nharm_lb[24] <- quantile(temp3, 0.025)
dat$nharm_ub[24] <- quantile(temp3, 0.975)

dat$harmrate_lb[24] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[24] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[24] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[24] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[25] <- quantile(temp3, 0.025)
dat$nharm_ub[25] <- quantile(temp3, 0.975)

dat$harmrate_lb[25] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[25] <- quantile(temp3 / temp1, 0.975)
```

```
dat$peharmrate_lb[25] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[25] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[26] <- quantile(temp3, 0.025)
dat$nharm_ub[26] <- quantile(temp3, 0.975)

dat$harmrate_lb[26] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[26] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[26] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[26] <- quantile(temp3 / temp2, 0.975)



dat$nharm_lb[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.025)
dat$nharm_ub[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.975)
```

```
dat$nharm_lb[28] <- quantile(temp3 / pharm_big3, 0.025)
dat$nharm_ub[28] <- quantile(temp3 / pharm_big3, 0.975)



dat$peharmrate_lb[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                        (temp2 / perror_big3 * (1 - perror_big3)), 0.025)
dat$peharmrate_ub[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                        (temp2 / perror_big3 * (1 - perror_big3)), 0.975)

dat$peharmrate_lb[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.025)
dat$peharmrate_ub[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.975)

dat$weight <- dat$peharmrate_point / peh_general
dat <- dat[, c(1:2, 9:10, 23, 3:5,
          6:8, 12:14, 20:22,
          11, 15, 16, 17:19)]

all(dat$misrate_lb < dat$misrate_point, na.rm = T)
all(dat$misrate_ub > dat$misrate_point, na.rm = T)
all(dat$harmrate_lb < dat$harmrate_point, na.rm = T)
all(dat$harmrate_ub > dat$harmrate_point, na.rm = T)
all(dat$nharm_lb < dat$nharm_point, na.rm = T)
all(dat$nharm_ub > dat$nharm_point, na.rm = T)
all(dat$nerror_lb < dat$nerror, na.rm = T)
all(dat$nerror_ub > dat$nerror, na.rm = T)

write.csv(dat, file = "big3_results_combo_updated.csv")
```

Supplemental File #1 for ***Burden of Serious Harms from Diagnostic Error in the USA***
by Newman-Toker, et al. BMJ Quality & Safety, 2023 (doi:10.1136/bmjqs-2021-014130)

Enclosed below are the main Supplementary Materials (Supplement 1) to the peer-reviewed, scientific journal publication entitled ***Burden of Serious Harms from Diagnostic Error in the USA*** *(2023).* Some of the methods descriptions (particularly in Supplement 1, section A5 about Monte Carlo analysis) are very similar to methods descriptions from the related, previously published manuscripts from earlier project phases. This is unavoidable since the statistical methods for the current manuscript were the same. An ancillary appendix (Supplement 2) provides the full statistical code for the Monte Carlo analysis.

Table of Contents for Supplement 1

Abbreviations Not Necessarily Defined in the Text

N/n – number
M – million
B – billion
CI – confidence interval
PR – plausible range
PPR – probabilistic plausible range

A) Methodological Details

A1. Study Methods – Rationale for Overall Architecture

Seeking valid estimates of disability and death from diagnostic error is important for public policy, yet methodologically challenging.[1] Diagnostic errors will necessarily be more frequent than diagnostic adverse events (of any severity), which, in turn, will be more frequent than serious misdiagnosis-related harms (i.e., permanent disability or death). Key concerns in estimation of serious harms[1] include (1) proper quantitative synthesis of the literature on error and harm rates, rather than extrapolations based on single studies conducted in non-representative settings; (2) judgments about attributable harms (i.e., the extent to which diagnostic errors result in serious harms or are potentially confounded by comorbidity); (3) judgments about preventability, including whether prevention will result in meaningful gains in healthy life years; (4) methodological risks of undercounting or overcounting serious harms, including double counting of deaths in patients who suffer more than one error; and (5) applying error and harm rate estimates to the appropriate population at risk and conducting methodologically robust statistical analyses to account for uncertainty associated with relatively low frequency events (i.e., serious harms).

Taking a disease-specific (rather than disease-agnostic) approach to measurement helps address many of these methodological challenges. It is easier to more rigorously and precisely measure diagnostic errors, harms, or preventability in disease-specific than disease-agnostic fashion since research studies that cut across diseases cannot incorporate rigorous reference standards for diagnostic accuracy or error for every possible condition. Synthesis of multiple studies via systematic review with meta-analysis is also more straightforward since disease-specific studies are more homogeneous in their disease and error definitions. Finally, a consistent finding from the literature on diagnostic error, whether derived from malpractice claims or clinical practice, is that vascular events, infections, and cancers (together known as the "Big Three" dangerous-disease categories) are responsible for three-quarters of serious misdiagnosis-related harms.[2,3] This permits extrapolation from these specific disease categories to all diseases.

As noted in the main manuscript Methods section, the overall study was designed to estimate the total annual burden (incidence) of serious harms from diagnostic error in the US. It was conducted in three study phases (Figure S1): (1) identify top diseases misdiagnosed that cause serious misdiagnosis-related harms (from a large, nationally representative malpractice data set previously coded for claim type, outcome/harm severity, and disease then comparing the proportion of "Big Three" diseases to clinical practice-based [non-claims] studies)[2]; (2) find disease-specific diagnostic error and harm rates for top harm-causing diseases (from clinical literature and vetted by experts)[4,5]; (3a) measure annual population incidence for each disease (from public use, nationally representative data sets) and (3b) combine error and harm rates with incidence to estimate total annual US incidence of serious diagnostic errors and harms. The first two study phases were published previously[2,4]; the final phase (3a, 3b) is included here.

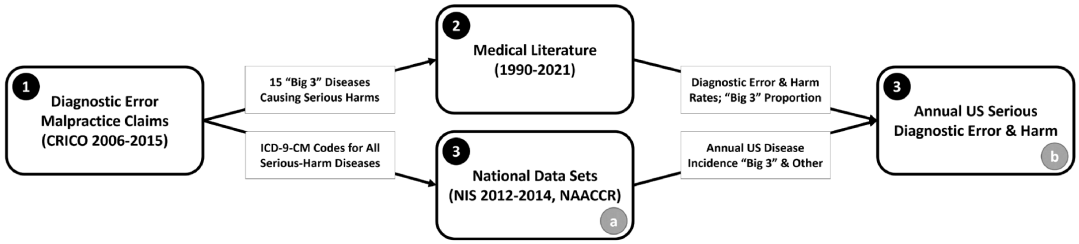

Figure S1. Overview of scientific methods for estimating total diagnostic errors and serious misdiagnosis-related harms. *Abbreviations: CRICO – Controlled Risk Insurance Company, Ltd.; ICD-9-CM – International Classification of Diseases 9th Revision, Clinical Modification; NAACCR – North American Association of Central Cancer Registries; NIS – National Inpatient Sample.*

A2. Role of Malpractice Claims in Estimation and Independence of Grand Total Harm Estimates

Some readers may wonder how it can be that our study began with a first phase that used malpractice claims data, yet the final estimates are somehow independent? As noted in the main manuscript Methods section, our scientific approach was constructed such that the final grand total estimates for errors and harms in the US are based on clinical literature and US population incidence, not malpractice claims. This is because (a) no error or harm rates were taken from claims-based studies, (b) the extrapolation from specific "Big Three" disease estimates to the grand total were based on the proportion of "Big Three" diseases causing errors and harms from clinical studies, and (c) any impact of having used malpractice claims to construct the original disease list or weights are mathematically unrelated to the grand totals.

Why were malpractice data used in the first place? Malpractice data were used to construct the initial list of diseases likely to be responsible for the greatest numbers of total misdiagnosis-related harms. We needed a starter list of such diseases in order to be able to take the disease-based approach (A1 above).

If malpractice data were good enough to make the disease list, why not just stop there? First, it is known that many medical errors never lead to malpractice claims, so it is hard to extrapolate even from a representative, national claims database to a true national estimate. In the Harvard Medical Practice Study, which compared negligent medical errors to malpractice claims, the chance that an injury caused by medical negligence would result in litigation was just 1.5% (95% CI 0-3.2).[6] Second, malpractice claims data are known to represent a biased sample. Some forms of bias in malpractice claims are well known, while others may be hidden. The most well-known bias in malpractice claims data is towards higher-severity harms. This is not necessarily a problem for estimating serious harms[7] (as in this study), but there may also be maldistributions of claims (i.e., non-representativeness) based on other factors as well. For example, myocardial infarction is probably overrepresented in claims relative to stroke as a cause of misdiagnosis-related harms—disease incidence is similar, diagnostic error rates are ~10-fold higher for stroke, and disabling neurologic injuries result in the highest claims payouts (so are more likely to spark a claim), yet numbers of claims are only ~1.5-fold higher for stroke; this could be because legal "standard of care" expectations for accuracy of heart attack diagnoses are higher than those for stroke.[5]

Were malpractice data used for any other purpose in constructing the estimates? Yes, as noted in the main manuscript Methods section, misdiagnosis-related harm rates were derived by combining high-quality data on disease-agnostic (non-disease specific) harms per diagnostic error from well-respected clinical studies then applying disease-specific harm-severity weights from malpractice claims.[4] A disease-agnostic approach was required because there were not a sufficient number of disease-specific studies examining attributable harm rates. We weighted the disease-agnostic, per-diagnostic-error serious harm rate for each disease to get a more realistic estimate of harms (e.g., aortic dissection is more likely lethal than pneumonia when initially missed, so assigning the same risk of serious harms *per diagnostic error* for each of the two diseases would have been inappropriate). For each disease, we multiplied the disease-specific, clinical literature-based diagnostic error rate ***by*** the clinical literature-based disease-agnostic per-error harm rate ***by*** a disease-specific, claims-based harm-severity weight. This weight was based on the disease-specific proportion of malpractice cases resulting in serious vs. non-serious harms (e.g., higher weight for aortic dissection than pneumonia). The weighting procedure was also used to prevent overcounting of harms from "other" (non-top 5) diseases. Full statistical details of this approach can be found in our prior publication's Supplementary materials (*Supplement A2, Requirements R1 and R4*[4]).

If that is true, then how can the final total estimates be claims-independent? The final results are independent of malpractice claims because we mathematically "forced" the proportion of errors and serious harms attributable to all combined Big Three diseases to be equal to the known attributable fractions found in the clinical literature (*see our prior publication's Supplement A2, Requirements R2 and*

R3[4]). This is described, in brief, in the Methods section of the main manuscript, "…"Big Three" results were used to calculate a grand total (including non-"Big Three" dangerous diseases) using the clinical proportion of diagnostic errors (58.5%) and serious harms (75.8%) attributable to "Big Three" diseases.[2,3] These proportions derive exclusively from research studies based in clinical practice (i.e., not malpractice claims studies) (*see Table 3 from our prior citation*[2]). Mathematically, the grand total of diagnostic errors was calculated by dividing the "Big Three" total number of diagnostic errors by 0.585. Similarly, the grand total of serious misdiagnosis-related harms was calculated by dividing the "Big Three" total number of serious misdiagnosis-related harms by 0.758." This forces independence from claims.

Then what are the implications of malpractice-claims based intermediate steps for the results? There are two main potential impacts of these claims-based steps. First, it is possible that the lower-ranked "top 5" diseases might be over-ranked (e.g., it is possible that the unnamed 6th-ranked disease categorized in the "other" subcategory in one of the "Big Three" categories might actually be the *real* 5th-ranked disease in that category). For example, the 6th-ranked disease in the infection category in malpractice claims was appendicitis. There were more than twice as many claims for endocarditis as appendicitis, which is why we searched out data on error rates for endocarditis rather than appendicitis. However, if malpractice claims were somehow biased towards endocarditis or away from appendicitis, it is still potentially conceivable that appendicitis might outrank endocarditis as a cause of misdiagnosis-related harms in clinical practice, since appendicitis has more than twice the real-world incidence of endocarditis. However, this is unlikely, because endocarditis is initially missed an estimated ~26% of the time[4] and appendicitis is initially missed no more than ~5% of the time, more than compensating for the higher incidence of appendicitis.[5] Note that, in our final analysis, appendicitis is still accounted for in the "other" infections subcategory (so it has not gone uncounted). Second (and related), it is possible that the relative proportion of "other" (non-top 5) diseases are underrepresented relative to the top 5.

In summary, serious harms estimated for individual diseases named (or unnamed) in Table 1 of the main manuscript are potentially impacted by unknown biases that could be present in malpractice claims. This could impact disease-specific rankings or the proportion of "other" (non-top 5) harms. However, the grand total harm estimates are mathematically fully independent of malpractice claims.

A3. Double Check of HCUP CCS Code Level Groupings Prior to NIS Incidence Analysis

Prior to NIS analysis of disease incidence in this third phase of the project, we performed a final cross-check at the code level using the HCUP CCS Level 3 groupings for vascular diseases and infections derived from the claims analysis.[2] The code lists were reviewed and any ICD codes unrelated to new, acute events (e.g., 438 "late effects of cerebrovascular disease") were removed prior to NIS analysis. We also reviewed all codes in the CRICO CBS data set to address issues of coding migration over time and reduce the risk that any specific codes might be missed because of sampling error in CRICO data during the years of analysis. From the wider code list, we found 18 related codes that belonged in the top 5 groupings (e.g., 433.0 "occlusion and stenosis of the basilar artery" and 435.3 "vertebrobasilar artery syndrome" for stroke) and added these before conducting the final NIS analysis. We did not consider Level 3 codes present in the parent HCUP CCS classification but *not* found in the malpractice claims data, to avoid any risk of overcounting non-life-or-limb-threatening diseases unlikely to cause harms. NIS analysis was run at both the ICD-9-CM code level as well as rolled up by disease and category to both (a) ensure sensibility and coherence and (b) identify any coding errors or gaps before being finalized.

A4. NIS Sampling & Weighting Procedures to Derive Nationally Representative Estimates

We followed standard procedures for NIS data to derive nationally representative estimates, which use pre-specified discharge weights to convert an unweighted sample of hospital discharges into a weighted,

nationally representative sample.[8] The result is a weighted estimate for both disease incidence and patient demographics. Each year in the US there are roughly 36M inpatient hospitalizations[9] at more than 6,000 hospitals.[10] For each year since 2012, NIS has sampled more than 7M hospital discharge records from more than 4,000 acute-care hospitals (excluding long-term acute care hospitals). The discharge weights are calculated by NIS data curators by first stratifying the NIS hospitals on the same variables that were used for creating the sample. These variables are hospital Census division, urban/rural location, teaching status, bed size, and ownership. A weight is then calculated for each stratum, by dividing the number of universe discharges (i.e., all discharges) in that stratum, obtained from HCUP and American Hospital Association data, by the number of NIS discharges (i.e., sampled discharges) in the stratum. Discharge weights are assigned to each sampled discharge by NIS data curators and are stored in the NIS data set for use in constructing nationally representative estimates. When discharge weights are applied to the unweighted NIS data, the result is an estimate of the number of discharges for the entire universe (i.e., an estimate of all acute care hospitalizations in the US).

A5. Monte Carlo Analysis to Determine Probabilistic Plausible Ranges (PPRs) *(reported previously[4])*

The main outcome measures were estimates of total US annual diagnostic errors and serious misdiagnosis-related harms. Annual incidence from NIS and NAACCR were multiplied by literature-derived estimates of disease-specific and category diagnostic error and harm rates,[4] an approach analogous to "minimal modeling" methods in cost-effectiveness or value-of-information analysis.[11]

To obtain the variability of these combined estimates, we used a probabilistic sampling approach based on Monte Carlo simulations[12] (Supplement 2). These simulations produce statistically valid 95% CIs that account for variability in both number and sample sizes for each disease. In the current manuscript, most of these uncertainty estimates are denoted as "probabilistic plausible ranges" (PPRs), rather than 95% CIs. This is because they rely, in part, on diagnostic error rates that utilize literature-derived (and expert-validated) plausible ranges (n=5 cancers) rather than statistically derived 95% CIs as their uncertainty range, reflecting some uncertainty beyond mere sampling error.[4] Specifically, experts felt that for the top five cancers, PRs should be wider than the statistical 95% CIs. For each cancer, this was because different studies defined diagnostic delays of different lengths—defining shorter delays as errors created an upper PR bound, while defining longer delays created a lower PR bound.[4] As part of the same Monte Carlo simulations, we also calculated PPRs around error and harm point estimates for the "other" (non-top 5) subcategories and combined categories (e.g., top 5 vascular events, total "Big Three," grand totals).

For the Monte Carlo analysis, skew-normal distributions were used to approximate the distributions of disease incidence rate, diagnostic error rate, and serious misdiagnosis-related harm rate, separately for each quantity. The location parameter of the skew-normal was set to be the point estimate of the corresponding rate. The scale and skewness parameters were determined such that the lower and upper bounds of 95% CI of the resulting skew-normal distribution coincided with each of the 95% CI or probabilistic range bounds of the corresponding rate. Due to the extreme skewness of plausible ranges for some diagnostic error rates,[4] all approximations were performed on logit-transformed distributions.[13] Monte Carlo samples were drawn independently from the resulting distributions. The population affected by the diagnostic error and the subsequent misdiagnosis-related harm were calculated for each Monte Carlo replica. The PPRs were given by the 2.5% and 97.5% quantiles based on 10,000,000 simulations. The large number of simulations was used to ensure tail probability and reduce Monte Carlo error due to the very skewed sampling distributions. Rates and other parameters have been published previously,[4] although diagnostic error rate estimates for stroke, myocardial infarction, venous thromboembolism, aortic aneurysm and dissection, and sepsis were updated to reflect the most robust estimates available from a systematic review with meta-analysis that was conducted by members of the authorship team.[5]

B) Estimated Impact of Methodological Choices & Assumptions on Results

B1. Sensitivity Analysis of Errors & Harms (Impact of Uncertainty in Model Parameters)

We conducted a one-way sensitivity analysis to assess the impact of parameter uncertainty on the final point estimates (Figure S2). Harm results were most sensitive to parameters for common infections. The three highest leverage parameters for potentially overestimating serious harms were the (1) other infection harm rate, (2) sepsis harm rate, and (3) pneumonia harm rate. However, even if each of these harm rates (and the one for stroke [4th for overestimates]) were placed at the lower plausible bound of harms assessed in this one-way sensitivity analysis, the point estimate of serious harms would still be over 500,000.

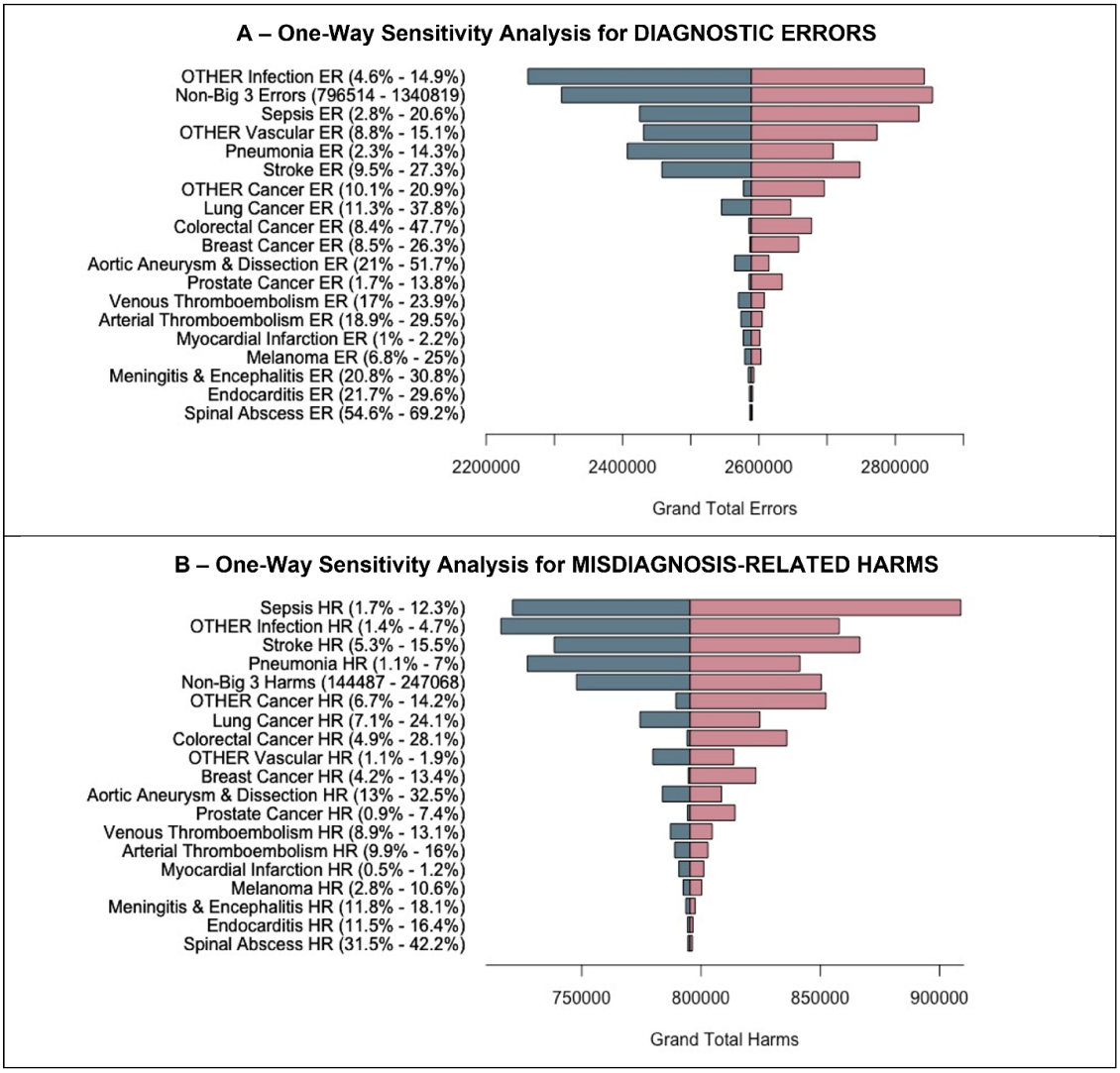

Figure S2. Sensitivity analysis of model parameters on diagnostic errors and serious harms. This one-way (as opposed to multi-way, probabilistic sensitivity analysis) varies one parameter value at a time while holding all other model parameter values constant. For each parameter, its value is ranged between the lower and upper uncertainty bounds (i.e., confidence interval or plausible range [*shown*]) and the impact of uncertainty on the total is illustrated, with highest impact parameters shown at the top and lowest at the bottom, yielding a "tornado" appearance. The

point estimate value used for each parameter is fixed as the "midline" of the tornado. The impact on the final total of using the lower bound parameter value is shown in ***blue to the left*** (reflecting possible ***overestimation*** in the point estimate). Conversely, the impact on the final total of using the upper bound parameter value is shown in ***red to the right*** (reflecting possible ***underestimation*** in the point estimate). Panel A shows a tornado diagram for diagnostic errors. The three parameters with the greatest potential for overestimation of errors were the (1) other infection error rate, (2) non-"Big Three" error rate, and (3) pneumonia error rate. Panel B shows a tornado diagram for serious harms. The three parameters with the greatest potential for overestimation of harms were the (1) other infection harm rate, (2) sepsis harm rate, and (3) pneumonia harm rate. *Abbreviations: ER – error rate; HR – harm rate.*

B2. Sensitivity Analysis of Errors & Harms (Impact of Using Only Principal Diagnosis)

We estimated disease incidence for vascular events and infections from HCUP data from the NIS. In the primary analysis, we counted discharge (or in-hospital death) diagnoses coded in either the principal or first-listed secondary diagnosis positions, as these two diagnoses are often of equal, competing weight.[14]

We conducted a sensitivity analysis using only first-position NIS diagnosis codes to assess the impact of this methodological decision on the final results. Primary vs. sensitivity results are shown in Table S1. Overall estimates of errors and harms were about 30% lower when using only principal diagnosis codes.

Table S1. Comparison of diagnostic errors and harms in the primary analysis vs. principal-only analysis*

Category	Primary Analysis N in thousands (PPR)	Principal Diagnosis-Only N in thousands (PPR)
Big Three Total Diagnostic Errors	1,514 (1,122-1,889)	1,044 (852-1,365)
Big Three Total Serious Harms	603 (454-776)	416 (344-550)
Grand Total Diagnostic Errors	2,588 (1,919-3,230)	1,785 (1,457-2,335)
Grand Total Serious Harms	795 (598-1,023)	549 (454-725)

* The primary analysis counted NIS diagnosis codes in either the principal or first-listed secondary positions. The sensitivity analysis counted NIS diagnosis codes in only the principal position (so are necessarily lower).

B3. Estimated Impact of Undercounting (Conservative Assumption re: Out-of-Hospital Deaths)

The conservative assumption was made that incident cases of dangerous (life or limb-threatening) vascular events and infections in the US would eventually involve a hospitalization, even if the patient was initially misdiagnosed in an ambulatory care setting. Outpatient (e.g., primary care, emergency department) visit diagnoses were not included separately in the disease incidence calculations because they would risk inflating disease incidence estimates through double counting. For example, had "myocardial infarction" cases that were correctly diagnosed in outpatient care (and then later confirmed as an inpatient) been included in the analysis, the same incident cases would be counted twice.

This methodological decision, chosen to avoid overcounting, does risk some degree of undercounting; this is principally via patients seen in outpatient settings (e.g., primary care or emergency department) who are misdiagnosed and then die in the community of their underlying illness, without ever reaching the hospital as an inpatient (i.e., out-of-hospital deaths following a missed diagnosis). Reliable data are not available to directly measure out-of-hospital deaths across all conditions (because death certificate data may be unreliable[15]). However, we can estimate the incidence of such deaths using pre-hospital death data, which are available for some of the more common dangerous diseases. For myocardial infarction, the proportion of all true cases that result in pre-hospital death was estimated in a rigorous population-based study from Germany to be 13.6%.[16] For stroke, we can estimate the rate by combining data on pre-hospital stroke deaths with data on stroke hospitalizations. In 2014 (the reference year for our study) there were 980 stroke hospitalizations per 100,000 among Medicare beneficiaries (who are predominantly patients aged 65 and older) in the state of New York.[17] With a 2014 New York state population aged 65

and older of 2,898,094,[18] the total number of stroke hospitalizations for those aged 65 and older in 2014 was 28,401. Since ~38% of stroke hospitalizations in 2014 occurred in patients below age 65,[19] the total number of New York state stroke hospitalizations in 2014 was approximately ~45,800. In 2012-2014, the crude rate of out-of-hospital stroke deaths in New York state was 11.6 per 100,000.[20] With a 2014 New York state total population of ~19,750,000,[18] the total number of pre-hospital stroke deaths was about ~2,300. Therefore, the proportion of strokes that resulted in death without making it to the hospital was about 4.8% of all strokes. The heart attack and stroke estimates cohere well with what is known about the natural history of these two diseases—harms from myocardial infarction are disproportionately deaths, while harms from stroke are disproportionately disabilities. Accordingly, it is sensible that a higher proportion of myocardial infarctions would result in pre-hospital deaths. If we take a disease incidence-weighted average of these two numbers (13.6% for heart attack and 4.8% for stroke), the estimated proportion of undercounted incident cases is ~10%.

If we postulate a similar overall rate (~10% out-of-hospital deaths) for the remaining vascular events and infections, and if 100% of these out-of-hospital death cases were deemed "misdiagnosed," our overall estimate of serious misdiagnosis-related harms would be under-counted by approximately 8% (since this particular problem is not likely to impact incident cancers, which are measured by different means of estimation). However, many such cases involve patients who die without ever having had the opportunity to be misdiagnosed because they never reached medical attention (e.g., previously asymptomatic sudden cardiac death from myocardial infarction[21] or strokes in which premonitory transient ischemic attack symptoms do not prompt the individual to seek attention[22]) (Figure S3). ***Thus, 8% represents a likely upper bound on undercounting and the true value is probably less than 5%***.

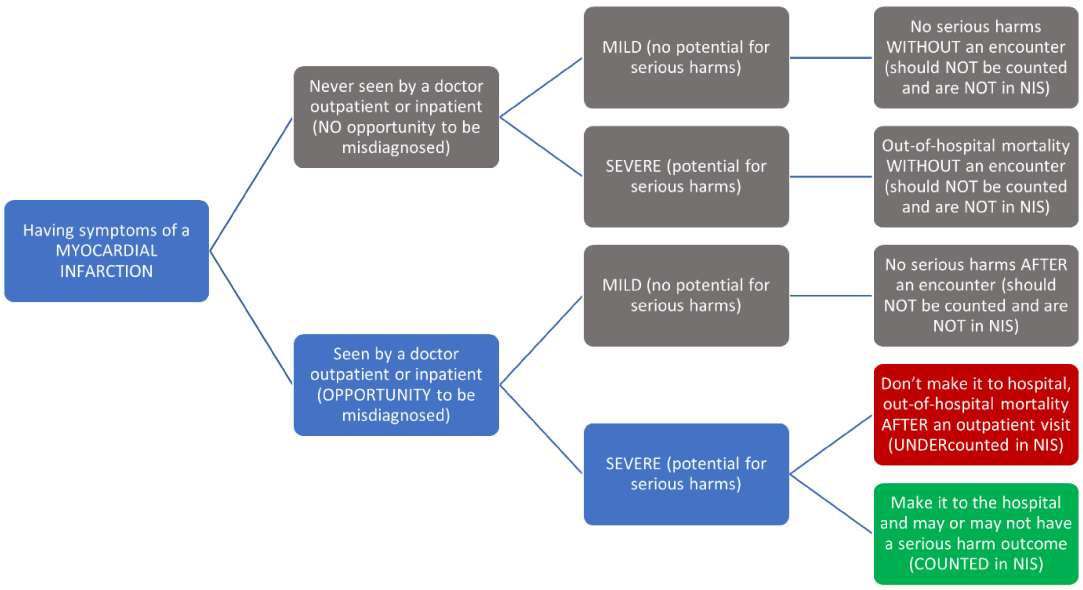

Figure S3. Conceptual relationship between disease, care-seeking behavior, and risk of undercounting. The figure illustrates how a specific subset of undercounted cases not in NIS data (shown in red) might impact the overall estimates of serious misdiagnosis-related harms; other undercounted cases would not (shown in gray). As described in the text, it is estimated that the impact of these undercounted cases is likely less than 5% of the total.
Abbreviation: NIS – National Inpatient Sample

B4. Estimated Impact of Overcounting (Based on Patients with More than One Hospitalization)

There is some possibility that our method of using inpatient hospitalizations to measure dangerous disease incidence might lead to overcounting. NIS data track hospitalizations, not patients, so some patients could have been admitted more than once (e.g., admitted for a myocardial infarction and later a stroke in the same year). Although a single person could suffer permanent disability in more than one way, one patient cannot die twice, so this could theoretically lead to overestimates of deaths using our method.

It is not possible to estimate the impact of such potential overcounting directly using NIS data, but we can estimate the potential extent of the problem by combining NIS with other data sources, such as the National Health Interview Survey (NHIS). Using NHIS, the Centers for Disease Control and Prevention (CDC) reports that, in 2014, ~17.5M (corresponding to ~76% of patients hospitalized at least once that year) were hospitalized only once, ~3.2M (corresponding to ~14% of patients hospitalized at least once that year) were hospitalized twice, and ~2.3M (corresponding to ~10% of patients hospitalized at least once that year) were hospitalized three or more times.[23] With ~35.4M total hospitalizations that year in NIS,[24] that means ~50% of hospitalizations involved "repeat visitors." Using NIS data from our current analysis (average for 2012-2014), 34% of hospitalizations were for vascular events or infections (i.e., 12.1M of 35.8M total hospitalizations). The clinical proportion of serious misdiagnosis-related harms represented by deaths is 46.7%.[25,26] Thus, the potential impact on overcounting deaths from missed vascular events and infections is ~8% (i.e., ~50% x ~34% x ~47% = ~8%). However, patients who did, in fact, die of a misdiagnosis from one of these illnesses could not have been counted again past their death date (i.e., they could not have been a "repeat visitor"), ***so the true value is likely to be lower (e.g., <5%)***.

C) Additional Validity Arguments

C1. Comparison with Independent Estimates from Diagnostic Errors in Hospital Autopsies

To gauge the plausibility of our overall serious misdiagnosis-related harms estimate (~795,000), we can derive the misdiagnosis-associated mortality in our data and compare it to that found in hospital autopsy data. We estimate total deaths from our current study using the previously published proportion of harms representing deaths across inpatient and outpatient settings—46.7%[25,26] (~795,000 x 46.7% = ~371,000). We can do the same for the principal-only analysis (~549,000 x 46.7% = ~256,000). Table S2 shows total expected US deaths and all serious harms (death plus permanent disability) in 2014 (our year of analysis), depending on the hypothesized proportion of deaths associated with diagnostic error.

Table S2. Anticipated US deaths and serious harms due to diagnostic error depending on hypothesized risk

Hypothetical Proportion of Deaths Associated with Diagnostic Error	Potential Misdiagnosis-Related Deaths in the US in 2014*	Corresponding Misdiagnosis-Related Serious Harms in 2014†
5%	131,321	281,027
10%	262,642	562,053
15%	393,963	843,080
20%	525,284	1,124,107

***** According to the Centers for Disease Control and Prevention (CDC), there were 2,626,418 (2.6M) US deaths in 2014,[27] which is the reference year chosen for our manuscript's analysis of serious misdiagnosis-related harms.
† Calculated using the clinical proportion of serious misdiagnosis-related harms represented by deaths (46.7%).[25,26]

The primary analysis estimate (~371,000 deaths [of ~2.6M deaths]) would represent a 14.1% (10.6-18.2) overall misdiagnosis-associated mortality nationally. The principal-only analysis (~256,000 deaths [of ~2.6M deaths]) would represent a 9.8% (8.1% to 12.9%) overall misdiagnosis-associated mortality

nationally. Either rate is higher than estimates from studies of hospital autopsies that consider epoch of diagnosis and adjust for bias from submaximal autopsy rates. A large meta-analysis of hospital autopsy studies projected that a modern US hospital which autopsied 100% of in-hospital deaths would find 8.4% (95% CI 5.2-13.1) suffered a major diagnostic error, half considered Class I (deaths directly attributed to the diagnostic error) and half Class II (diagnostic errors that would have changed clinical management and could have altered the patient's clinical course).[28]

However, it is expected that the population-based proportion of misdiagnosis-related deaths would be higher than that found in hospital autopsies. One reason is that hospital autopsies consider diagnostic errors not yet recognized at the time of death, but usually not errors occurring pre-hospitalization when a prompt intervention might have been lifesaving. For example, consider a patient with a new, abrupt-onset headache who is sent home from a primary care clinic as "migraine." If the patient were to return a week later to the emergency department in a coma, they might be promptly diagnosed with aneurysmal subarachnoid hemorrhage and admitted to the intensive care unit. Were they to die, a hospital autopsy would then indicate that no diagnostic error had occurred *in the hospital*. The risk of death from brain aneurysm is increased nearly 5-fold after an initial misdiagnosis, and misdiagnosis disproportionately occurs in outpatient clinics with isolated headache clinical presentations.[29,30] Given highly effective treatments for brain aneurysm and the knowledge that prognosis post-operatively is almost entirely tied to clinical severity at the time of surgery,[29] this case should clearly count as a potentially preventable death due to diagnostic error, but would be considered a correct ante-mortem diagnosis in hospital autopsy data.

Another reason is that the proportion of deaths associated with diagnostic error/delay is probably higher for out-of-hospital than in-hospital deaths. Estimated diagnostic error rates in primary care (2.4% per visit [n=5,126/212,165][31,32]) exceed those in hospitals (0.7%[33]). This makes sense, since (a) hospitalization tends to occur relatively late in the natural course of illness, when a patient has become sick enough to merit inpatient care, and often after the underlying cause for their symptoms is more obvious, and (b) hospitalized patients undergo more intensive diagnostic testing and monitoring than ambulatory patients. As a result, it would not be surprising if outpatient deaths were more often pursuant to diagnostic errors. Since >80% of healthcare visits occur in non-ED ambulatory care[34,35] and >65% of all US deaths occur outside the hospital,[36] an overall misdiagnosis-associated mortality of 9.8% to 14.1% seems plausible.

How many of these misdiagnosis-associated deaths are preventable and how much (or little) longevity might potentially be reclaimed for affected patients is uncertain.[37] Nevertheless, individual cases of otherwise healthy young patients who die from treatable causes that were misdiagnosed make it clear that this could be a half-century or more in years of quality life lost for a given patient.[38-40] The same is true for lifelong disability in young patients after missed opportunities to promptly treat disabling diseases.[41-43] For some of the most harmful diseases in our list, correct initial diagnosis has been associated with clear and substantial reductions in morbidity or mortality (e.g., ischemic stroke [~5-fold],[5] aneurysmal subarachnoid hemorrhage [~5-fold],[29] and ruptured abdominal aortic aneurysm [~2-fold][44]).

C2. Comparison with Independent Estimates from Diagnostic Adverse Events in Hospitals

We can also gauge the plausibility of our serious harm results in light of diagnostic adverse event data from inpatient hospital stays. Gunderson et al. recently published a systematic review of hospital-based studies of diagnostic adverse events (n=22), two of these US-based.[33] They estimated a pooled hospital misdiagnosis-related harm rate (counting any harm severity) of 0.7% (95% CI 0.5-1.1) with high levels of heterogeneity (I^2=95%, p<0.001) (overall range across studies 0.1-2.7).[33] Most of these studies did not report specific diseases missed, but eight did (n=136 cases). Authors listed 70 diseases or categories with at least two instances (Table 2[33]). Among these 70, 78.6% were attributed to "Big Three" diseases (this distribution is very similar to the attributable % used in our current population-based study [75.8%[2]]).

If applied to US-based hospitals, they estimated ~250,000 patients harmed annually from diagnostic error. They were unable to assess harm severity based on the available literature. In the well-designed 2010 Dutch study by Zwaan et al.,[25] which measured a similar rate of hospital-based diagnostic adverse events (0.4%), they found that, of diagnostic adverse events, 29.1% resulted in death and ~25.6% (estimated from their Figure 1) resulted in disability at discharge, for an overall rate of ~54.7% serious harms.[25]

Using NIS 2014 US hospitalizations (~35.4M),[24] this translates to ~135,000 (~97K-213K) (uncertainty estimated using 95% CI from Gunderson[33]) serious misdiagnosis-related harms in US hospitals. Relative to our 2014 primary analysis estimate of ~795,000 suffering death or permanent disability, this suggests that ~17% (~12-27) of serious misdiagnosis-related harms occur among inpatients and ~83% (~73-88) among outpatients. Although in 2014 only 2.7% of the roughly 1.3B US healthcare visits were inpatient hospitalizations,[4] severity of illness and diagnostic error adverse events are both higher than outpatient.[45] To help gauge this effect, the proportion of high-severity misdiagnosis-related harms linked to inpatient care in malpractice claims is ~28% (i.e., >10-fold over-representation relative to visit proportion).[2] It is reasonable to expect that inpatient malpractice claims for diagnostic adverse events would be artificially over-represented in claims relative to clinical care proportions, since, relative to outpatient care, outcome severity is higher[45] (a known predictor of legal action[7]) and a "paper trail" of documentation to establish a legal action is more readily available (another likely predictor[2]). Therefore, the ~28% represents an "upper bound," of sorts, on the inpatient-attributable serious harms fraction. Thus, an estimate that ~17% (~12-27) of US serious misdiagnosis-related harms occur in inpatient settings seems quite plausible.

Using these numbers, we can also estimate that the total annual hospital-based deaths from diagnostic error in the US in 2014 would be ~35.4M (NIS 2014[24]) x 0.7% (95% CI 0.5-1.1) (Gunderson et al.[33]) x 29.1% (Zwaan et al.[25]) = ~72,000 (~51K-113K). This value is squarely within the range projected by Leape, Berwick, and Bates (i.e., 40,000-80,000 hospital deaths per year)[46] derived by multiplying total hospital deaths by the rate of hospital autopsy-determined diagnostic errors. The ~72,000 misdiagnosis-related hospital deaths estimate is also squarely in the range of what is expected based on a rigorous systematic review of hospital autopsies by Shojania et al.[28] They calculated the combined Goldman Class I/II diagnostic error rate for an average, modern, US-based hospital that autopsied 100% of its deaths—8.4% (95% CI 5.2-13.1).[28] According to the CDC, there were 2,626,418 (2.6M) US deaths in 2014,[27] of which 37.3% were hospital-based,[47] for a total of ~980,000 hospital deaths. The ~72,000 would therefore correspond to a 7.4% (n=~72,000/~980,000) misdiagnosis-attributable fraction of hospital deaths.

The resulting estimates comparing inpatient-only harms to those across settings are shown in Table S3.

Table S3. US deaths and serious harms due to diagnostic error in 2014 comparing inpatient to all settings

Misdiagnosis-Related Harms	Inpatient Only (Prior Studies[25,33])	Across Settings (Current Study)
Total Serious (Death + Disability)	~135,000 (~97K-213K)*	~795,000 (~598K-1,023K)†
Deaths Only	~72,000 (~51K-113K)*	~371,000 (~279K-478K)‡
Disability (*calculated difference*)	~63,000 (~46K-100K)	~424,000 (~319-545)

* Uncertainty accounted for using 95% CI from Gunderson[33] plus serious harms or death % from Zwaan.[25]
† Uncertainty accounted for using PPR from primary analysis, which used Monte Carlo simulations (see A5).
‡ Uncertainty accounted for using PPR from primary analysis and point death % from Zwaan[25] & Singh[26] combined.

C3. Triangulation of Available Data across Sources and Methods

We have described three separate methods of estimation that all yield compatible results:

- Method 1 (Manuscript): disease incidence x literature-based misdiagnosis-related harm rate
- Method 2 (Section C1): hospital deaths x % of deaths attributable to diagnostic error
- Method 3 (Section C2): hospital adverse events x % of adverse events resulting in harm or death

The three distinct methods can be used to derive inpatient serious harms and misdiagnosis-related deaths. Point estimates for total inpatient serious harms across the three methods range from ~135,000-225,000. All three methods produce point estimates for deaths that fall within the tight range of ~72,000-105,000.

Method 1 gives us total serious misdiagnosis-related harms (i.e., death + disability) in 2014 for inpatient and outpatient settings (~795,000). From this, we can estimate total deaths using the previously published proportion of harms representing deaths across inpatient and outpatient settings—46.7%[25,26] (~371,000) (see C1). Combining this with the proportion of total serious harms attributed to inpatient settings from a large, nationally representative sample of malpractice claims (28%[2]) gives ~225,000 total serious harms and ~105,000 deaths in US hospitals annually. Because of likely bias towards legal action for inpatient claims (see C2), these are presumed to be slight overestimates. Despite this, they are still close to results estimated by Methods 2 and 3, below. Method 1 serious harms (~225,000) fall within the uncertainty range by Method 2 (~96K-241K) and just beyond that by Method 3 (~97K-213K). Method 1 deaths (~105,000) fall within the uncertainty ranges by Method 2 (~51K-128K) and Method 3 (~51K-113K).

Method 2 gives us total misdiagnosis-related hospital deaths directly. Hospital deaths in 2014 (~980,000 [see C2]) were published by the CDC and the misdiagnosis-attributable fraction (8.4% [95% CI 5.2-13.1]) is from a rigorous meta-analysis of 53 autopsy studies whose final estimates account for study country, study epoch, and submaximal autopsy rate.[28] The estimate is ~82,000 (~51K-128K) misdiagnosis-related hospital inpatient deaths. Using the previously published proportion of serious harms representing deaths in the inpatient setting (~53.2% [29.1% deaths of ~54.7% serious harms][25]), we can estimate total serious harms of ~155,000 (~96K-241K). These are close to results from Method 3, despite different derivations.

Method 3 gives us total misdiagnosis-related hospital harms or deaths. US hospitalizations in 2014 (~35.4M[24]) derive from the NIS. The diagnostic adverse event rate (0.7%) is from a meta-analysis of 22 studies of hospital-based diagnostic adverse events,[33] and the proportion of adverse events resulting in serious harms (~54.7%) or death (29.1%) are from a rigorous, population-based sample of inpatient diagnostic adverse events from 21 hospitals. The resulting hospital estimate is ~135,000 (~97K-213K) serious misdiagnosis-related harms, which includes an estimated ~72,000 deaths (~51K-113K) (see C2).

Thus, our current results triangulate well across data sources and methods (convergent construct validity). When the consistency of these estimated misdiagnosis-related harms is combined with the consistent proportion of serious harms (~75-80%[3]) accounted for by "Big Three" diseases across settings (n=44 studies in primary care, emergency department, hospital[3]), this enhances the validity of our study results.

Finally, our results are bolstered by coherence with another recent systematic review and meta-analysis of diagnostic errors in the emergency department,[5] which permits a rough estimate across clinical settings as a final cross-check. As shown in Table S4, data combined from other sources (~855,000 [plausible range ~490K-1,659K]) align well with those found in the current study (~795,000 [PPR 598K-1,023K]). Both values translate to a US per-healthcare-visit serious misdiagnosis-related harm rate of about 0.08%.

Table S4. Serious misdiagnosis-related harms and serious harms rates by clinical setting (alternate sources)

Clinical Setting in which Error Occurs	Annual US Visits per Year in 2014 (n)	Total Serious Misdiagnosis-Related Harms (n)	Estimated Serious Harms Rate (%)
Inpatient	35,400,000[24]	~135,000 (~97K-213K)[25,33]*	~0.38% (~0.27-0.60)*
Emergency Department	137,800,000[48]	~430,000† (~259K-1,042K)†	~0.31% (~0.19-0.76)[5]
Primary Care Clinics	461,800,000[49]	~206,000† (~103K-309K)‡	~0.04%§ (~0.02-0.07)‡
Specialty Care Clinics	423,000,000[49]	~85,000† (~42K-127K)‡	~0.02%§ (~0.01-0.03)‡
TOTAL (£)	1,057,900,000	~855,000 (~490K-1,659K)	~0.08% (~0.05-0.16)

* Point estimate and uncertainty combine US inpatient harms estimate from Gunderson[33] with serious harms proportion from Zwaan[25] as described in Section C3 above. Harms rate is then calculated using visits per year.

† Total serious misdiagnosis-related harms calculated as serious harms rate x annual US visits.

‡ When precise estimates of uncertainty were lacking, we arbitrarily assigned it as +/- 50% of the point estimate.

§ Although the rate of serious harms from diagnostic error in ambulatory clinic-based care is not well characterized, generally the risks of a serious harm event (on a per visit basis) are much lower than either inpatient or emergency department care, simply because the severity of illness is much lower. From data in patients with missed stroke (erroneously called "benign" dizziness), we can estimate that the risks of serious harms after discharge from primary care are approximately 7-fold lower than those seen in the emergency department[50]; we can also approximate that the rate of serious harms after discharge from specialty care is likely about half that seen in primary care.[51]

£ Totals were calculated by summing n's for visits and harms, then dividing harms by visits to get rates (weighted average). As a result, the lower and upper uncertainty bounds are wider than if they were sampled probabilistically.

D) Supplementary References *(N.B. – citation numbers differ from main manuscript)*

1. Shojania KG, Dixon-Woods M. Estimating deaths due to medical error: the ongoing controversy and why it matters. *BMJ Qual Saf.* 2017;26(5):423-428.

2. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis (Berl).* 2019;6(3):227-240.

3. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Corrigendum to: Serious misdiagnosis-related harms in malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis (Berl).* 2020;8(1):127-128.

4. 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).* 2020;8(1):67-84.

5. Newman-Toker DE, Peterson SM, Badihian S, Hassoon A, Nassery N, Parizadeh D, Wilson LM, Jia Y, Omron R, Tharmarajah S, Guerin L, Bastani PB, Fracica EA, Kotwal S, Robinson KA. *Diagnostic Errors in the Emergency Department: A Systematic Review*. Comparative Effectiveness Review No. 258. (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No. 75Q80120D00003.) AHRQ Publication No. 22(23)-EHC043. Rockville, MD: Agency for Healthcare Research and Quality; December 2022 (addendum in press, anticipated August 2023). DOI: https://doi.org/10.23970/AHRQEPCCER258.

6. Localio AR, Lawthers AG, Brennan TA, et al. Relation between malpractice claims and adverse events due to negligence. Results of the Harvard Medical Practice Study III. *N Engl J Med.* 1991;325(4):245-251.

7. Studdert DM, Mello MM, Gawande AA, et al. Claims, errors, and compensation payments in medical malpractice litigation. *N Engl J Med.* 2006;354(19):2024-2033.

8. Producing National HCUP Estimates - Accessible Version. Agency for Healthcare Research and Quality, 2015. (Accessed May 20, 2023 at https://www.hcup-us.ahrq.gov/tech_assist/nationalestimates/508_course/508course_2018.jsp.)

9. National Trends in Inpatient Stays. Healthcare Cost and Utilization Project (HCUP) Fastats. Agency for Healthcare Research & Quality, 2023. (Accessed May 20, 2023 at https://datatools.ahrq.gov/hcup-fast-stats; https://datatools.ahrq.gov/sites/default/files/2022-10/DownloadTable_NationalTrendsInpatientStays_10-3-22.xls.)

10. Fast Facts on U.S. Hospitals, 2022. American Hospital Association. (Accessed May 20, 2023 at https://www.aha.org/statistics/fast-facts-us-hospitals.)

11. Meltzer DO, Hoomans T, Chung JW, Basu A. Minimal modeling approaches to value of information analysis for health research. *Med Decis Making.* 2011;31(6):E1-E22.

12. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* 1970;57(1):97–109.

13. Bartlett MS. The use of transformations. *Biometrics.* 1947;3(1):39-52.

14. ICD-10-CM Official Guidelines for Coding and Reporting FY 2019. Centers for Medicare & Medicaid Services, 2019. (Accessed May 20, 2023 at https://www.cms.gov/Medicare/Coding/ICD10/Downloads/2019-ICD10-Coding-Guidelines-.pdf.)

15. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One.* 2014;9(3):e92286.

16. Amann U, Heier M, Thilo C, Linseisen J, Meisinger C. Determinants of prehospital coronary heart disease death. *Sci Rep.* 2021;11(1):17134.

17. Interactive Atlas of Heart Disease and Stroke. Centers for Disease Control and Prevention. (Accessed May 20, 2023 at https://nccd.cdc.gov/DHDSPAtlas/reports.aspx?geographyType=county&state=NY&themeId=20&filterIds=1,5,3,6,7&filterOptions=6,1,1,1,1#report.)

18. Annual Report of Vital Statistics: New York State 2014. New York State Department of Health. (Accessed May 20, 2023 at https://www.health.ny.gov/statistics/vital_statistics/docs/vital_statistics_annual_report_2014.pdf.)

19. Stroke Facts. Centers for Disease Control and Prevention. (Accessed May 20, 2023 at https://www.cdc.gov/stroke/facts.htm.)

20. Wales K. Pre-Transport Stroke Death Rates by EMS Region and Location of Stroke Designated Hospitals, New York State. Centers for Disease Control and Prevention. (Accessed May 20, 2023 at https://www.cdc.gov/dhdsp/maps/gisx/mapgallery/NY-stroke.html.)

21. O'Neill L. Estimating out-of-hospital mortality due to myocardial infarction. *Health Care Manag Sci.* 2003;6(3):147-154.

22. Giles MF, Flossman E, Rothwell PM. Patient behavior immediately after transient ischemic attack according to clinical characteristics, perception of the event, and predicted risk of stroke. *Stroke.* 2006;37(5):1254-1260.

23. Summary Health Statistics: National Health Interview Survey, 2014. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2014_SHS_Table_P-10.pdf.)

24. Trends in Hospital Inpatient Stays in the United States, 2005-2014 (HCUP Statistical Brief #225). Agency for Healthcare Research & Quality. (Accessed May 20, 2023 at https://www.hcup-us.ahrq.gov/reports/statbriefs/sb225-Inpatient-US-Stays-Trends.jsp.)

25. Zwaan L, de Bruijne M, Wagner C, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med.* 2010;170(12):1015-1021.

26. 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(6):418-425.

27. Deaths: Final Data for 2014. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://www.cdc.gov/nchs/data/dvs/LCWK1_2014.pdf.)

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

29. Kowalski RG, Claassen J, Kreiter KT, et al. Initial misdiagnosis and outcome after subarachnoid hemorrhage. *JAMA.* 2004;291(7):866-869.

30. 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(15):1468-1477.

31. Singh H, Giardina TD, Forjuoh SN, et al. Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ Qual Saf.* 2012;21(2):93-100.

32. 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(9):727-731.

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

34. Ambulatory Care Use and Physician Office Visits. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at http://www.cdc.gov/nchs/fastats/physician-visits.htm.)

35. Emergency Department Visits. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at http://www.cdc.gov/nchs/fastats/emergency-department.htm.)

36. Trends in Inpatient Hospital Deaths: National Hospital Discharge Survey, 2000–2010. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://www.cdc.gov/nchs/products/databriefs/db118.htm.)

37. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA.* 2001;286(4):415-420.

38. From a Small Scrape to Sepsis - Rory Staunton's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/from-a-small-scrape-to-sepsis/.)

39. Unbelieved - Jessica Barnett's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/unbelieved-jessica-barnetts-story/.)

40. Tested Too Late - Chad Becken's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/tested-too-late-chad-beckens-story/.)

41. Missed Stroke Diagnosis - John Michael Night's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/missed-stroke-diagnosis/.)

42. The Missed Bilirubin Test - Cal Sheridan's Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/the-missed-bilirubin-test-cal-sheridans-story/.)

43. The Lingering Lump - Diana Cejas' Story. Society to Improve Diagnosis in Medicine. (Accessed May 20, 2023 at https://www.improvediagnosis.org/stories_posts/the-lingering-lump-diana-cejas-story/.)

44. Smidfelt K, Nordanstig J, Davidsson A, Torngren K, Langenskiold M. Misdiagnosis of ruptured abdominal aortic aneurysms is common and is associated with increased mortality. *J Vasc Surg.* 2021;73(2):476-483 e473.

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

46. Leape LL, Berwick DM, Bates DW. Counting deaths due to medical errors (in reply). *JAMA.* 2002;288(19):2404-2405.

47. QuickStats: Percentage Distribution of Deaths, by Place of Death — United States, 2000–2014. *Morbidity and Mortality Weekly Report.* 2016;65:357. (Accessed May 20, 2023 at https://www.cdc.gov/mmwr/volumes/65/wr/mm6513a6.htm.)

48. Hooker EA, Mallow PJ, Oglesby MM. Characteristics and Trends of Emergency Department Visits in the United States (2010-2014). *J Emerg Med.* 2019;56(3):344-351.

49. National Ambulatory Medical Care Survey: 2014 State and National Summary Tables. Centers for Disease Control and Prevention, National Center for Health Statistics. (Accessed May 20, 2023 at https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2014_namcs_web_tables.pdf.)

50. Nassery N, Mane K, Liu F, et al. Measuring missed strokes using administrative and claims data: towards a diagnostic performance dashboard to monitor diagnostic errors. Diagnostic Error in Medicine 2016; November 6-8, 2016, 2016; Hollywood, CA.

51. Chang TP, Bery AK, Wang Z, et al. Stroke hospitalization after misdiagnosis of "benign dizziness" is lower in specialty care than general practice: a population-based cohort analysis of missed stroke using SPADE methods. *Diagnosis (Berl).* 2021;9(1):96-106.

Supplemental File #2 for ***Burden of Serious Harms from Diagnostic Error in the USA***
by Newman-Toker, et al. BMJ Quality & Safety, 2023 (doi:10.1136/bmjqs-2021-014130)

Enclosed below are additional Supplementary Materials (Supplement 2) to the peer-reviewed, scientific journal publication entitled ***Burden of Serious Harms from Diagnostic Error in the USA*** *(2023).*

This is the full statistical code (R v4.2.2, Vienna, Austria) for the Monte Carlo analysis.

```
# load packages
rm(list = ls())
library(sn)
library(rootSolve)
library(gdata)
library(MASS)
library(DEoptim)

set.seed(37)
logit <- function(x) {
    out <- log(x/(1-x))
    return(out)
}

expit <- function(x) {
    out <- exp(x) / (1 + exp(x))
    return(out)
}

#### primary analysis  ####
dat <- read.csv(file = "big3_data_dx1.csv")

peh_general <- 374/1216 # general per-error-harm rate
```

```
peh_general_se <- sqrt(1/374 + 1/(1216-374)) # standard error on logit scale
peh_general_lower <- expit(logit(peh_general) + qnorm(0.025) * peh_general_se)
peh_general_upper <- expit(logit(peh_general) + qnorm(0.975) * peh_general_se)

perror_big3 <- 717 / (717 + 509) # proportion of big 3 in errors
pharm_big3 <- 157 / 207 # proportion of big 3 in harms

vascular_index <- which(dat$category == "vascular" & dat$disease != "OTHER Vascular")
vascular_other_index <- which(dat$disease == "OTHER Vascular")
infection_index <- which(dat$category == "infection" & dat$disease != "OTHER Infection")
infection_other_index <- which(dat$disease == "OTHER Infection")
cancer_index <- which(dat$category == "cancer" & dat$disease != "OTHER Cancer")
cancer_other_index <- which(dat$disease == "OTHER Cancer")

dat$nerror <- dat$incidence_point * dat$misrate_point
shp <- dat$proportion_highharm # severity harm proportion

peh_fun <- function(peh_general, nerror) {
 # function to calculate disease-specific per-error-harm rates
   peh_rate <- rep(NA, nrow(dat))
   peh_rate[c(vascular_index, infection_index, cancer_index)] <- peh_general * pharm_big3 /
perror_big3 *
       (sum(nerror) * shp[c(vascular_index, infection_index, cancer_index)]) /
       (sum(nerror[vascular_index] * shp[vascular_index]) * 1684 / 1344 +
           sum(nerror[infection_index] * shp[infection_index]) * 992 / 600 +
           sum(nerror[cancer_index] * shp[cancer_index]) * 2793 / 1529)
   peh_rate[vascular_other_index] <- sum(nerror[vascular_index] * peh_rate[vascular_index] *
334/1334) /
       nerror[vascular_other_index]
```

```
    peh_rate[infection_other_index] <- sum(nerror[infection_index] * peh_rate[infection_index] *
392/600) /
        nerror[infection_other_index]
    peh_rate[cancer_other_index] <- sum(nerror[cancer_index] * peh_rate[cancer_index] * 1264/1529)
/
        nerror[cancer_other_index]
    return(peh_rate)
}

# point estimate of per-error-serious harm rates
dat$peharmrate_point <- peh_fun(peh_general, dat$nerror)
dat$peharmrate_lb <- peh_fun(peh_general_lower, dat$nerror)
dat$peharmrate_ub <- peh_fun(peh_general_upper, dat$nerror)

# Monte Carlo simulation, to construct plausible intervals for harm rate and number of harm
snest <- function(u, lower, upper, lowerq = 0.025, upperq = 0.975) {
    solfun <- function(beta) {
        omega <- beta[1]
        alpha <- beta[2]
        out1 <- psn(lower, xi = u, omega = omega, alpha = alpha) - lowerq
        out2 <- psn(upper, xi = u, omega = omega, alpha = alpha) - upperq
        return((out1^2+out2^2))
    }
    start <- c(((upper-lower)/4)^2, 0) + mvrnorm(100, mu = rep(0, 2), Sigma = diag(1, 2))
    start[, 1] <- abs(start[, 1])
    tempfit <- t(apply(start, 1, function(x) {
        fit <- optim(fn = solfun, par = x,
                control = list(maxit = 500000))
        return(c(fit$par, fit$value))
```

```
    }))
    tempfit <- tempfit[complete.cases(tempfit), ]
    return(tempfit[which(tempfit[, 3] == min(tempfit[, 3]))[1], 1:3])
}



totl <- 316283434
nmc <- 1*(10^7)
numerical_error <- NULL
dat$nerror_lb <- NA
dat$nerror_ub <- NA
dat$nharm_point <- dat$nerror * dat$peharmrate_point
dat$nharm_lb <- NA
dat$nharm_ub <- NA
dat$harmrate_point <- dat$misrate_point * dat$peharmrate_point
dat$harmrate_lb <- NA
dat$harmrate_ub <- NA

# Monte Carlo simulation based on skew-normal estimation
mc_result <- list()
for (i in 1:nrow(dat)) {
    mc_result[[i]] <- list()
    if (dat$incidence_lb[i] == dat$incidence_ub[i]) {
        mc_incidence <- rep(dat$incidence_point[i], nmc)
    } else {
        fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                               lower = logit(dat$incidence_lb[i] / totl),
                               upper = logit(dat$incidence_ub[i] / totl))
        while(fit_incidence[3] > 10^(-10)) {
```

```
            fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                                   lower = logit(dat$incidence_lb[i] / totl),
                                   upper = logit(dat$incidence_ub[i] / totl))
        }
        mc_incidence <- expit(rsn(n = nmc, xi = logit(dat$incidence_point[i] / totl),
                                  omega = fit_incidence[1], alpha = fit_incidence[2])) * totl
    }
    temp <- (c(quantile(mc_incidence, c(0.025, 0.975)) - c(dat$incidence_lb[i], dat$incidence_ub[i]))) /
totl

    if (i %in% c(vascular_index, infection_index, cancer_index)) {
        fit_misrate <- snest(logit(dat$misrate_point[i]),
                             lower = logit(dat$misrate_lb[i]),
                             upper = logit(dat$misrate_ub[i]))
        while(fit_misrate[3] > 10^(-10)) {
            fit_misrate <- snest(logit(dat$misrate_point[i]),
                                 lower = logit(dat$misrate_lb[i]),
                                 upper = logit(dat$misrate_ub[i]))
        }

        mc_misrate <- expit(rsn(n = nmc, xi = logit(dat$misrate_point[i]),
                                omega = fit_misrate[1], alpha = fit_misrate[2]))
        temp <- c(temp, quantile(mc_misrate, c(0.025, 0.975)) - c(dat$misrate_lb[i], dat$misrate_ub[i]))

        mc_result[[i]]$fit_misrate <- fit_misrate
        mc_result[[i]]$mc_misrate <- mc_misrate
    } else {
        temp <- c(temp, rep(NA, 2))
    }
```

```
    fit_peh <- snest(logit(dat$peharmrate_point[i]),
                     lower = logit(dat$peharmrate_lb[i]),
                     upper = logit(dat$peharmrate_ub[i]))
    while(fit_peh[3] > 10^(-10)) {
        fit_peh <- snest(logit(dat$peharmrate_point[i]),
                         lower = logit(dat$peharmrate_lb[i]),
                         upper = logit(dat$peharmrate_ub[i]))
    }
    mc_peh <- expit(rsn(n = nmc, xi = logit(dat$peharmrate_point[i]),
                        omega = fit_peh[1], alpha = fit_peh[2]))
    temp <- c(temp, quantile(mc_peh, c(0.025, 0.975)) - c(dat$peharmrate_lb[i], dat$peharmrate_ub[i]))

    numerical_error <- rbind(numerical_error, temp)

    mc_result[[i]]$fit_incidence <- fit_incidence
    mc_result[[i]]$mc_incidence <- mc_incidence

    mc_result[[i]]$fit_peh <- fit_peh
    mc_result[[i]]$mc_peh <- mc_peh
    mc_result[[i]]$numerical_error <- numerical_error
}

summary(numerical_error)

# calculate plausible intervals for "OTHER" vascular/infection/censor harm rate (misrate)
# because point estimate is calculated as weighted sum and plausible intervals should be calcualted accordingly
temp1 <- 0
```

```
temp2 <- 0
for (i in vascular_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[vascular_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate,
0.025)
dat$misrate_ub[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate,
0.975)



temp1 <- 0
temp2 <- 0
for (i in infection_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[infection_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate,
0.025)
dat$misrate_ub[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate,
0.975)


temp1 <- 0
temp2 <- 0
for (i in cancer_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
```

```
mc_result[[cancer_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.975)



# calculate plausible intervals for number of errors, number of harms and harm rate
# based on Monte Carlo generated data
for (i in 1:nrow(dat)) {
    mc_incidence <- mc_result[[i]]$mc_incidence
    mc_misrate <- mc_result[[i]]$mc_misrate
    mc_peh <- mc_result[[i]]$mc_peh
    dat$nerror_lb[i] <- quantile(mc_incidence * mc_misrate, 0.025)
    dat$nerror_ub[i] <- quantile(mc_incidence * mc_misrate, 0.975)

    dat$nharm_lb[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.025)
    dat$nharm_ub[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.975)

    dat$harmrate_lb[i] <- quantile(mc_misrate * mc_peh, 0.025)
    dat$harmrate_ub[i] <- quantile(mc_misrate * mc_peh, 0.975)
}

# calculate subtotals and plausible intervals
# create structure for subtotals
temp <- c("Top 5 Vascular Subtotal",
        "Top 5 Infection Subtotal",
        "Top 5 Cancer Subtotal",
        "Top 15 Subtotal",
        "Total Vascular",
```

```
        "Total Infection",
        "Total Cancer",
        "Big 3 Subtotal",
        "All Other Dx Errors",
        "Grand Total")
temp2 <- data.frame(matrix(NA, nrow = length(temp), ncol = ncol(dat)))
colnames(temp2) <- colnames(dat)
temp2[, 2] <- temp
temp2[, 1] <- rep("Subtotals", length(temp))
dat <- rbind(dat, temp2)

dat[, -c(1:2)] <- apply(dat[, -c(1:2)], c(1, 2), as.numeric)

# calculate subtotal incidence
dat$incidence_point[19] <- sum(dat$incidence_point[1:5])
dat$incidence_point[20] <- sum(dat$incidence_point[7:11])
dat$incidence_point[21] <- sum(dat$incidence_point[13:17])
dat$incidence_point[22] <- sum(dat$incidence_point[19:21])

dat$incidence_point[23] <- sum(dat$incidence_point[1:6])
dat$incidence_point[24] <- sum(dat$incidence_point[7:12])
dat$incidence_point[25] <- sum(dat$incidence_point[13:18])
dat$incidence_point[26] <- sum(dat$incidence_point[1:18])


temp <-lapply(1:18, function(i) return(mc_result[[i]]$mc_incidence))

temp1 <- rep(0, nmc)
for (i in 1:5) {
```

```
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[19] <- quantile(temp1, 0.025)
dat$incidence_ub[19] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[20] <- quantile(temp1, 0.025)
dat$incidence_ub[20] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[21] <- quantile(temp1, 0.025)
dat$incidence_ub[21] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[22] <- quantile(temp1, 0.025)
dat$incidence_ub[22] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
```

```
for (i in 1:6) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[23] <- quantile(temp1, 0.025)
dat$incidence_ub[23] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[24] <- quantile(temp1, 0.025)
dat$incidence_ub[24] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[25] <- quantile(temp1, 0.025)
dat$incidence_ub[25] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[26] <- quantile(temp1, 0.025)
dat$incidence_ub[26] <- quantile(temp1, 0.975)
```

```
dat[27:28, 3:5] <- NA

# calculate subtotal number of errors
dat$nerror[19] <- sum(dat$nerror[1:5])
dat$nerror[20] <- sum(dat$nerror[7:11])
dat$nerror[21] <- sum(dat$nerror[13:17])
dat$nerror[22] <- sum(dat$nerror[19:21])

dat$nerror[23] <- sum(dat$nerror[1:6])
dat$nerror[24] <- sum(dat$nerror[7:12])
dat$nerror[25] <- sum(dat$nerror[13:18])
dat$nerror[26] <- sum(dat$nerror[1:18])

dat$nerror[27] <- dat$nerror[26] / perror_big3 * (1 - perror_big3)
dat$nerror[28] <- dat$nerror[26] / perror_big3



temp1 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[19] <- quantile(temp1, 0.025)
dat$nerror_ub[19] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
```

```
}
dat$nerror_lb[20] <- quantile(temp1, 0.025)
dat$nerror_ub[20] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[21] <- quantile(temp1, 0.025)
dat$nerror_ub[21] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[22] <- quantile(temp1, 0.025)
dat$nerror_ub[22] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[23] <- quantile(temp1, 0.025)
dat$nerror_ub[23] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
```

```
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[24] <- quantile(temp1, 0.025)
dat$nerror_ub[24] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[25] <- quantile(temp1, 0.025)
dat$nerror_ub[25] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[26] <- quantile(temp1, 0.025)
dat$nerror_ub[26] <- quantile(temp1, 0.975)

dat$nerror_lb[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.025)
dat$nerror_ub[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.975)

dat$nerror_lb[28] <- quantile(temp1 / perror_big3, 0.025)
dat$nerror_ub[28] <- quantile(temp1 / perror_big3, 0.975)
```

```
# calculate subtotal misrate
dat$misrate_point[19:28] <- dat$nerror[19:28] / dat$incidence_point[19:28]

temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[19] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[19] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[20] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[20] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
```

```
dat$misrate_lb[21] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[21] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[22] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[22] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[23] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[23] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
```

```
}
dat$misrate_lb[24] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[24] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[25] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[25] <- quantile(temp1 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[26] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[26] <- quantile(temp1 / temp2, 0.975)



# calculate subtotal nharm, peh, and harm rate
dat$nharm_point[19] <- sum(dat$nharm_point[1:5])
dat$nharm_point[20] <- sum(dat$nharm_point[7:11])
```

```
dat$nharm_point[21] <- sum(dat$nharm_point[13:17])
dat$nharm_point[22] <- sum(dat$nharm_point[19:21])

dat$nharm_point[23] <- sum(dat$nharm_point[1:6])
dat$nharm_point[24] <- sum(dat$nharm_point[7:12])
dat$nharm_point[25] <- sum(dat$nharm_point[13:18])
dat$nharm_point[26] <- sum(dat$nharm_point[1:18])

dat$nharm_point[27] <- dat$nharm_point[26] / pharm_big3 * (1 - pharm_big3)
dat$nharm_point[28] <- dat$nharm_point[26] / pharm_big3



dat$harmrate_point[19:28] <- dat$nharm_point[19:28] / dat$incidence_point[19:28]
dat$peharmrate_point[19:28] <- dat$nharm_point[19:28] / dat$nerror[19:28]



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[19] <- quantile(temp3, 0.025)
dat$nharm_ub[19] <- quantile(temp3, 0.975)

dat$harmrate_lb[19] <- quantile(temp3 / temp1, 0.025)
```

```
dat$harmrate_ub[19] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[19] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[19] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[20] <- quantile(temp3, 0.025)
dat$nharm_ub[20] <- quantile(temp3, 0.975)

dat$harmrate_lb[20] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[20] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[20] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[20] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[21] <- quantile(temp3, 0.025)
dat$nharm_ub[21] <- quantile(temp3, 0.975)

dat$harmrate_lb[21] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[21] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[21] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[21] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
```

```
}

dat$nharm_lb[22] <- quantile(temp3, 0.025)
dat$nharm_ub[22] <- quantile(temp3, 0.975)

dat$harmrate_lb[22] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[22] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[22] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[22] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[23] <- quantile(temp3, 0.025)
dat$nharm_ub[23] <- quantile(temp3, 0.975)

dat$harmrate_lb[23] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[23] <- quantile(temp3 / temp1, 0.975)
```

```
dat$peharmrate_lb[23] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[23] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[24] <- quantile(temp3, 0.025)
dat$nharm_ub[24] <- quantile(temp3, 0.975)

dat$harmrate_lb[24] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[24] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[24] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[24] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
```

```
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[25] <- quantile(temp3, 0.025)
dat$nharm_ub[25] <- quantile(temp3, 0.975)

dat$harmrate_lb[25] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[25] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[25] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[25] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}
```

```
dat$nharm_lb[26] <- quantile(temp3, 0.025)
dat$nharm_ub[26] <- quantile(temp3, 0.975)

dat$harmrate_lb[26] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[26] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[26] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[26] <- quantile(temp3 / temp2, 0.975)



dat$nharm_lb[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.025)
dat$nharm_ub[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.975)

dat$nharm_lb[28] <- quantile(temp3 / pharm_big3, 0.025)
dat$nharm_ub[28] <- quantile(temp3 / pharm_big3, 0.975)



dat$peharmrate_lb[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                                  (temp2 / perror_big3 * (1 - perror_big3)), 0.025)
dat$peharmrate_ub[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                                  (temp2 / perror_big3 * (1 - perror_big3)), 0.975)

dat$peharmrate_lb[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.025)
dat$peharmrate_ub[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.975)

dat$weight <- dat$peharmrate_point / peh_general
dat <- dat[, c(1:2, 9:10, 23, 3:5,
        6:8, 12:14, 20:22,
        11, 15, 16, 17:19)]
```

```
all(dat$misrate_lb < dat$misrate_point, na.rm = T)
all(dat$misrate_ub > dat$misrate_point, na.rm = T)
all(dat$harmrate_lb < dat$harmrate_point, na.rm = T)
all(dat$harmrate_ub > dat$harmrate_point, na.rm = T)
all(dat$nharm_lb < dat$nharm_point, na.rm = T)
all(dat$nharm_ub > dat$nharm_point, na.rm = T)
all(dat$nerror_lb < dat$nerror, na.rm = T)
all(dat$nerror_ub > dat$nerror, na.rm = T)

write.csv(dat, file = "big3_results_dx1_updated.csv")



#### secondary analysis  ####
dat <- read.csv(file = "big3_data_combo.csv")
set.seed(37)
peh_general <- 374/1216 # general per-error-harm rate
peh_general_se <- sqrt(1/374 + 1/(1216-374)) # standard error on logit scale
peh_general_lower <- expit(logit(peh_general) + qnorm(0.025) * peh_general_se)
peh_general_upper <- expit(logit(peh_general) + qnorm(0.975) * peh_general_se)

perror_big3 <- 717 / (717 + 509) # big 3 error rate
pharm_big3 <- 157 / 207 # harm rate

vascular_index <- which(dat$category == "vascular" & dat$disease != "OTHER Vascular")
vascular_other_index <- which(dat$disease == "OTHER Vascular")
infection_index <- which(dat$category == "infection" & dat$disease != "OTHER Infection")
```

```
infection_other_index <- which(dat$disease == "OTHER Infection")

cancer_index <- which(dat$category == "cancer" & dat$disease != "OTHER Cancer")

cancer_other_index <- which(dat$disease == "OTHER Cancer")


dat$nerror <- dat$incidence_point * dat$misrate_point

shp <- dat$proportion_highharm # severity harm proportion


peh_fun <- function(peh_general, nerror) {

 # function to calculate disease-specific per-error-harm rates

    peh_rate <- rep(NA, nrow(dat))

    peh_rate[c(vascular_index, infection_index, cancer_index)] <- peh_general * pharm_big3 /
perror_big3 *

        (sum(nerror) * shp[c(vascular_index, infection_index, cancer_index)]) /

        (sum(nerror[vascular_index] * shp[vascular_index]) * 1684 / 1344 +

            sum(nerror[infection_index] * shp[infection_index]) * 992 / 600 +

            sum(nerror[cancer_index] * shp[cancer_index]) * 2793 / 1529)

    peh_rate[vascular_other_index] <- sum(nerror[vascular_index] * peh_rate[vascular_index] *
334/1334) /

        nerror[vascular_other_index]

    peh_rate[infection_other_index] <- sum(nerror[infection_index] * peh_rate[infection_index] *
392/600) /

        nerror[infection_other_index]

    peh_rate[cancer_other_index] <- sum(nerror[cancer_index] * peh_rate[cancer_index] * 1264/1529)
/

        nerror[cancer_other_index]

    return(peh_rate)

}


# point estimate of per-error-serious harm rates

dat$peharmrate_point <- peh_fun(peh_general, dat$nerror)
```

```
dat$peharmrate_lb <- peh_fun(peh_general_lower, dat$nerror)
dat$peharmrate_ub <- peh_fun(peh_general_upper, dat$nerror)

# Monte Carlo simulation, to construct plausible intervals for harm rate and number of harm
snest <- function(u, lower, upper, lowerq = 0.025, upperq = 0.975) {
    solfun <- function(beta) {
        omega <- beta[1]
        alpha <- beta[2]
        out1 <- psn(lower, xi = u, omega = omega, alpha = alpha) - lowerq
        out2 <- psn(upper, xi = u, omega = omega, alpha = alpha) - upperq
        return((out1^2+out2^2))
    }
    start <- c(((upper-lower)/4)^2, 0) + mvrnorm(100, mu = rep(0, 2), Sigma = diag(1, 2))
    start[, 1] <- abs(start[, 1])
    tempfit <- t(apply(start, 1, function(x) {
        fit <- optim(fn = solfun, par = x,
                control = list(maxit = 500000))
        return(c(fit$par, fit$value))
    }))
    tempfit <- tempfit[complete.cases(tempfit), ]
    return(tempfit[which(tempfit[, 3] == min(tempfit[, 3]))[1], 1:3])
}

totl <- 316283434
nmc <- 1*(10^7)
numerical_error <- NULL
dat$nerror_lb <- NA
dat$nerror_ub <- NA
dat$nharm_point <- dat$nerror * dat$peharmrate_point
```

```
dat$nharm_lb <- NA
dat$nharm_ub <- NA
dat$harmrate_point <- dat$misrate_point * dat$peharmrate_point
dat$harmrate_lb <- NA
dat$harmrate_ub <- NA

# Monte Carlo simulation based on skew-normal estimation
mc_result <- list()
for (i in 1:nrow(dat)) {
    mc_result[[i]] <- list()
    if (dat$incidence_lb[i] == dat$incidence_ub[i]) {
        mc_incidence <- rep(dat$incidence_point[i], nmc)
    } else {
        fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                               lower = logit(dat$incidence_lb[i] / totl),
                               upper = logit(dat$incidence_ub[i] / totl))
        while(fit_incidence[3] > 10^(-10)) {
            fit_incidence <- snest(logit(dat$incidence_point[i] / totl),
                                   lower = logit(dat$incidence_lb[i] / totl),
                                   upper = logit(dat$incidence_ub[i] / totl))
        }
        mc_incidence <- expit(rsn(n = nmc, xi = logit(dat$incidence_point[i] / totl),
                                  omega = fit_incidence[1], alpha = fit_incidence[2])) * totl
    }
    temp <- (c(quantile(mc_incidence, c(0.025, 0.975)) - c(dat$incidence_lb[i], dat$incidence_ub[i]))) /
totl

    if (i %in% c(vascular_index, infection_index, cancer_index)) {
        fit_misrate <- snest(logit(dat$misrate_point[i]),
```

```
                    lower = logit(dat$misrate_lb[i]),
                    upper = logit(dat$misrate_ub[i]))
    while(fit_misrate[3] > 10^(-10)) {
        fit_misrate <- snest(logit(dat$misrate_point[i]),
                    lower = logit(dat$misrate_lb[i]),
                    upper = logit(dat$misrate_ub[i]))
    }

    mc_misrate <- expit(rsn(n = nmc, xi = logit(dat$misrate_point[i]),
                    omega = fit_misrate[1], alpha = fit_misrate[2]))
    temp <- c(temp, quantile(mc_misrate, c(0.025, 0.975)) - c(dat$misrate_lb[i], dat$misrate_ub[i]))

    mc_result[[i]]$fit_misrate <- fit_misrate
    mc_result[[i]]$mc_misrate <- mc_misrate
} else {
    temp <- c(temp, rep(NA, 2))
}

fit_peh <- snest(logit(dat$peharmrate_point[i]),
            lower = logit(dat$peharmrate_lb[i]),
            upper = logit(dat$peharmrate_ub[i]))
while(fit_peh[3] > 10^(-10)) {
    fit_peh <- snest(logit(dat$peharmrate_point[i]),
                lower = logit(dat$peharmrate_lb[i]),
                upper = logit(dat$peharmrate_ub[i]))
}
mc_peh <- expit(rsn(n = nmc, xi = logit(dat$peharmrate_point[i]),
                omega = fit_peh[1], alpha = fit_peh[2]))
temp <- c(temp, quantile(mc_peh, c(0.025, 0.975)) - c(dat$peharmrate_lb[i], dat$peharmrate_ub[i]))
```

```
    numerical_error <- rbind(numerical_error, temp)

    mc_result[[i]]$fit_incidence <- fit_incidence
    mc_result[[i]]$mc_incidence <- mc_incidence

    mc_result[[i]]$fit_peh <- fit_peh
    mc_result[[i]]$mc_peh <- mc_peh
    mc_result[[i]]$numerical_error <- numerical_error
}
summary(numerical_error) # 10e-5

# calculate plausible intervals for "OTHER" vascular/infection/censor harm rate (misrate)
# because point estimate is calculated as weighted sum and plausible intervals should be calcualted accordingly
temp1 <- 0
temp2 <- 0
for (i in vascular_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[vascular_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[vascular_other_index] <- quantile(mc_result[[vascular_other_index]]$mc_misrate, 0.975)


temp1 <- 0
temp2 <- 0
```

```
for (i in infection_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[infection_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[infection_other_index] <- quantile(mc_result[[infection_other_index]]$mc_misrate, 0.975)

temp1 <- 0
temp2 <- 0
for (i in cancer_index) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
mc_result[[cancer_other_index]]$mc_misrate <- temp1 / temp2
dat$misrate_lb[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.025)
dat$misrate_ub[cancer_other_index] <- quantile(mc_result[[cancer_other_index]]$mc_misrate, 0.975)

# calculate plausible intervals for number of errors, number of harms and harm rate
# based on Monte Carlo generated data
for (i in 1:nrow(dat)) {
    mc_incidence <- mc_result[[i]]$mc_incidence
    mc_misrate <- mc_result[[i]]$mc_misrate
    mc_peh <- mc_result[[i]]$mc_peh
    dat$nerror_lb[i] <- quantile(mc_incidence * mc_misrate, 0.025)
```

```
    dat$nerror_ub[i] <- quantile(mc_incidence * mc_misrate, 0.975)

    dat$nharm_lb[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.025)
    dat$nharm_ub[i] <- quantile(mc_incidence * mc_misrate * mc_peh, 0.975)

    dat$harmrate_lb[i] <- quantile(mc_misrate * mc_peh, 0.025)
    dat$harmrate_ub[i] <- quantile(mc_misrate * mc_peh, 0.975)
}

# calculate subtotals and plausible intervals
# create structure for subtotals
temp <- c("Top 5 Vascular Subtotal",
        "Top 5 Infection Subtotal",
        "Top 5 Cancer Subtotal",
        "Top 15 Subtotal",
        "Total Vascular",
        "Total Infection",
        "Total Cancer",
        "Big 3 Subtotal",
        "All Other Dx Errors",
        "Grand Total")
temp2 <- data.frame(matrix(NA, nrow = length(temp), ncol = ncol(dat)))
colnames(temp2) <- colnames(dat)
temp2[, 2] <- temp
temp2[, 1] <- rep("Subtotals", length(temp))
dat <- rbind(dat, temp2)

dat[, -c(1:2)] <- apply(dat[, -c(1:2)], c(1, 2), as.numeric)
```

```
# calculate subtotal incidence
dat$incidence_point[19] <- sum(dat$incidence_point[1:5])
dat$incidence_point[20] <- sum(dat$incidence_point[7:11])
dat$incidence_point[21] <- sum(dat$incidence_point[13:17])
dat$incidence_point[22] <- sum(dat$incidence_point[19:21])

dat$incidence_point[23] <- sum(dat$incidence_point[1:6])
dat$incidence_point[24] <- sum(dat$incidence_point[7:12])
dat$incidence_point[25] <- sum(dat$incidence_point[13:18])
dat$incidence_point[26] <- sum(dat$incidence_point[1:18])



temp <-lapply(1:18, function(i) return(mc_result[[i]]$mc_incidence))

temp1 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[19] <- quantile(temp1, 0.025)
dat$incidence_ub[19] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[20] <- quantile(temp1, 0.025)
dat$incidence_ub[20] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
```

```
for (i in 13:17) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[21] <- quantile(temp1, 0.025)
dat$incidence_ub[21] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[22] <- quantile(temp1, 0.025)
dat$incidence_ub[22] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[23] <- quantile(temp1, 0.025)
dat$incidence_ub[23] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[24] <- quantile(temp1, 0.025)
dat$incidence_ub[24] <- quantile(temp1, 0.975)
```

```
temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[25] <- quantile(temp1, 0.025)
dat$incidence_ub[25] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + temp[[i]]
}
dat$incidence_lb[26] <- quantile(temp1, 0.025)
dat$incidence_ub[26] <- quantile(temp1, 0.975)

dat[27:28, 3:5] <- NA

# calculate subtotal number of errors
dat$nerror[19] <- sum(dat$nerror[1:5])
dat$nerror[20] <- sum(dat$nerror[7:11])
dat$nerror[21] <- sum(dat$nerror[13:17])
dat$nerror[22] <- sum(dat$nerror[19:21])

dat$nerror[23] <- sum(dat$nerror[1:6])
dat$nerror[24] <- sum(dat$nerror[7:12])
dat$nerror[25] <- sum(dat$nerror[13:18])
dat$nerror[26] <- sum(dat$nerror[1:18])
```

```
dat$nerror[27] <- dat$nerror[26] / perror_big3 * (1 - perror_big3)
dat$nerror[28] <- dat$nerror[26] / perror_big3



temp1 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[19] <- quantile(temp1, 0.025)
dat$nerror_ub[19] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[20] <- quantile(temp1, 0.025)
dat$nerror_ub[20] <- quantile(temp1, 0.975)



temp1 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[21] <- quantile(temp1, 0.025)
dat$nerror_ub[21] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
```

```
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[22] <- quantile(temp1, 0.025)
dat$nerror_ub[22] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[23] <- quantile(temp1, 0.025)
dat$nerror_ub[23] <- quantile(temp1, 0.975)


temp1 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[24] <- quantile(temp1, 0.025)
dat$nerror_ub[24] <- quantile(temp1, 0.975)

temp1 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[25] <- quantile(temp1, 0.025)
dat$nerror_ub[25] <- quantile(temp1, 0.975)
```

```
temp1 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
}
dat$nerror_lb[26] <- quantile(temp1, 0.025)
dat$nerror_ub[26] <- quantile(temp1, 0.975)

dat$nerror_lb[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.025)
dat$nerror_ub[27] <- quantile(temp1 / perror_big3 * (1 - perror_big3), 0.975)

dat$nerror_lb[28] <- quantile(temp1 / perror_big3, 0.025)
dat$nerror_ub[28] <- quantile(temp1 / perror_big3, 0.975)




# calculate subtotal misrate
dat$misrate_point[19:28] <- dat$nerror[19:28] / dat$incidence_point[19:28]

temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[19] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[19] <- quantile(temp1 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[20] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[20] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[21] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[21] <- quantile(temp1 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[22] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[22] <- quantile(temp1 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[23] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[23] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[24] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[24] <- quantile(temp1 / temp2, 0.975)


temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[25] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[25] <- quantile(temp1 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp2 <- temp2 + mc_result[[i]]$mc_incidence
}
dat$misrate_lb[26] <- quantile(temp1 / temp2, 0.025)
dat$misrate_ub[26] <- quantile(temp1 / temp2, 0.975)


# calculate subtotal nharm, peh, and harm rate
dat$nharm_point[19] <- sum(dat$nharm_point[1:5])
dat$nharm_point[20] <- sum(dat$nharm_point[7:11])
dat$nharm_point[21] <- sum(dat$nharm_point[13:17])
dat$nharm_point[22] <- sum(dat$nharm_point[19:21])

dat$nharm_point[23] <- sum(dat$nharm_point[1:6])
dat$nharm_point[24] <- sum(dat$nharm_point[7:12])
dat$nharm_point[25] <- sum(dat$nharm_point[13:18])
dat$nharm_point[26] <- sum(dat$nharm_point[1:18])

dat$nharm_point[27] <- dat$nharm_point[26] / pharm_big3 * (1 - pharm_big3)
dat$nharm_point[28] <- dat$nharm_point[26] / pharm_big3


dat$harmrate_point[19:28] <- dat$nharm_point[19:28] / dat$incidence_point[19:28]
```

```
dat$peharmrate_point[19:28] <- dat$nharm_point[19:28] / dat$nerror[19:28]



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:5) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[19] <- quantile(temp3, 0.025)
dat$nharm_ub[19] <- quantile(temp3, 0.975)

dat$harmrate_lb[19] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[19] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[19] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[19] <- quantile(temp3 / temp2, 0.975)



temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:11) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
```

```
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[20] <- quantile(temp3, 0.025)
dat$nharm_ub[20] <- quantile(temp3, 0.975)

dat$harmrate_lb[20] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[20] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[20] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[20] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:17) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[21] <- quantile(temp3, 0.025)
dat$nharm_ub[21] <- quantile(temp3, 0.975)

dat$harmrate_lb[21] <- quantile(temp3 / temp1, 0.025)
```

```
dat$harmrate_ub[21] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[21] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[21] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in c(1:5, 7:11, 13:17)) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[22] <- quantile(temp3, 0.025)
dat$nharm_ub[22] <- quantile(temp3, 0.975)

dat$harmrate_lb[22] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[22] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[22] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[22] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:6) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[23] <- quantile(temp3, 0.025)
dat$nharm_ub[23] <- quantile(temp3, 0.975)

dat$harmrate_lb[23] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[23] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[23] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[23] <- quantile(temp3 / temp2, 0.975)
```

```
temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 7:12) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
```

```
}

dat$nharm_lb[24] <- quantile(temp3, 0.025)
dat$nharm_ub[24] <- quantile(temp3, 0.975)

dat$harmrate_lb[24] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[24] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[24] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[24] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 13:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[25] <- quantile(temp3, 0.025)
dat$nharm_ub[25] <- quantile(temp3, 0.975)

dat$harmrate_lb[25] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[25] <- quantile(temp3 / temp1, 0.975)
```

```
dat$peharmrate_lb[25] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[25] <- quantile(temp3 / temp2, 0.975)




temp1 <- rep(0, nmc)
temp2 <- rep(0, nmc)
temp3 <- rep(0, nmc)
for (i in 1:18) {
    temp1 <- temp1 + mc_result[[i]]$mc_incidence
    temp2 <- temp2 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate
    temp3 <- temp3 + mc_result[[i]]$mc_incidence * mc_result[[i]]$mc_misrate * mc_result[[i]]$mc_peh
}

dat$nharm_lb[26] <- quantile(temp3, 0.025)
dat$nharm_ub[26] <- quantile(temp3, 0.975)

dat$harmrate_lb[26] <- quantile(temp3 / temp1, 0.025)
dat$harmrate_ub[26] <- quantile(temp3 / temp1, 0.975)

dat$peharmrate_lb[26] <- quantile(temp3 / temp2, 0.025)
dat$peharmrate_ub[26] <- quantile(temp3 / temp2, 0.975)



dat$nharm_lb[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.025)
dat$nharm_ub[27] <- quantile(temp3 / pharm_big3 * (1 - pharm_big3), 0.975)
```

```
dat$nharm_lb[28] <- quantile(temp3 / pharm_big3, 0.025)
dat$nharm_ub[28] <- quantile(temp3 / pharm_big3, 0.975)



dat$peharmrate_lb[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                        (temp2 / perror_big3 * (1 - perror_big3)), 0.025)
dat$peharmrate_ub[27] <- quantile((temp3 / pharm_big3 * (1 - pharm_big3)) /
                        (temp2 / perror_big3 * (1 - perror_big3)), 0.975)

dat$peharmrate_lb[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.025)
dat$peharmrate_ub[28] <- quantile((temp3 / pharm_big3) / (temp2 / perror_big3), 0.975)

dat$weight <- dat$peharmrate_point / peh_general
dat <- dat[, c(1:2, 9:10, 23, 3:5,
          6:8, 12:14, 20:22,
          11, 15, 16, 17:19)]

all(dat$misrate_lb < dat$misrate_point, na.rm = T)
all(dat$misrate_ub > dat$misrate_point, na.rm = T)
all(dat$harmrate_lb < dat$harmrate_point, na.rm = T)
all(dat$harmrate_ub > dat$harmrate_point, na.rm = T)
all(dat$nharm_lb < dat$nharm_point, na.rm = T)
all(dat$nharm_ub > dat$nharm_point, na.rm = T)
all(dat$nerror_lb < dat$nerror, na.rm = T)
all(dat$nerror_ub > dat$nerror, na.rm = T)

write.csv(dat, file = "big3_results_combo_updated.csv")
```