



Delft University of Technology

How safe is the diagnostic process in healthcare?

Marang-Van De Mheen, Perla J.; Thomas, Eric J.; Graber, Mark L.

DOI

[10.1136/bmjqs-2023-016496](https://doi.org/10.1136/bmjqs-2023-016496)

Publication date

2023

Document Version

Final published version

Published in

BMJ Quality and Safety

Citation (APA)

Marang-Van De Mheen, P. J., Thomas, E. J., & Graber, M. L. (2023). How safe is the diagnostic process in healthcare? *BMJ Quality and Safety*, 33(2), 82-85. Article bmjqs-2023-016496. <https://doi.org/10.1136/bmjqs-2023-016496>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.




Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

How safe is the diagnostic process in healthcare?

Perla J Marang-van de Mheen ^{1,2} Eric J Thomas ^{3,4}
Mark L Graber ⁵

¹Safety & Security Science, Delft University of Technology, Faculty of Technology, Policy & Management, Delft, The Netherlands

²Centre for Safety in Healthcare, Delft University of Technology, Delft, The Netherlands

³Internal Medicine, University of Texas John P and Katherine G McGovern Medical School, Houston, Texas, USA

⁴The UTHealth-Memorial Hermann Center for Healthcare Quality and Safety, UTHealth, Houston, Texas, USA

⁵Stony Brook University, Stony Brook, New York, USA

Correspondence to

Dr Perla J Marang-van de Mheen, Safety & Security Science, Delft University of Technology, Faculty of Technology, Policy & Management, Delft 2628 BX, The Netherlands; p.j.marang-vandemheen@tudelft.nl

Accepted 12 September 2023



► <http://dx.doi.org/10.1136/bmjqs-2021-014130>



© Author(s) (or their employer(s)) 2023. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Marang-van de Mheen PJ, Thomas EJ, Graber ML. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2023-016496

The seminal report *To Err is Human* focused on a wide range of serious patient safety concerns; diagnostic error was mentioned only in passing.¹ Very little data were available on the magnitude of harm related to diagnostic errors at that time, except for a back-of-the-napkin estimate that diagnostic error could be responsible for 40 000–80 000 in-hospital deaths annually.² The problem finally received its due 15 years later, when the National Academy of Medicine asserted that “... most of us will experience at least one diagnostic error in our lifetime, sometimes with devastating consequences”.³

In this context, the paper by Newman-Toker *et al* in this issue of *BMJ Quality & Safety* is a welcome contribution, presenting an extensively researched set of estimates that proposes that harm may be an order-of-magnitude larger.⁴ The paper is the third part of a larger study, building on two previous studies that (1) identified the top diseases causing serious misdiagnosis-related harms^{5 6} and (2) estimated disease-specific diagnostic error and harm rates for these top harm-causing diseases, based on literature review and expert verification.⁷ In this final third part, these results are combined with nationwide disease incidence data to generate extrapolations on the number of patients harmed by diagnostic error across care settings in the USA. The total serious harms annually in the USA were estimated to be 795 000 (plausible range 598 000–1 023 000), including 371 000 deaths and 424 000 disabilities. Sensitivity analyses using more conservative assumptions estimated 549 000 serious harms. The top five dangerous diseases (stroke, sepsis, pneumonia, venous thromboembolism and lung cancer) accounted for 38.7%. The authors conclude that harm associated with diagnostic error far surpasses any other patient safety concern, and

is probably the largest source of death across all care settings linked to medical error, while also acknowledging that it is uncertain how many serious harms can be prevented. To place these estimates in context, the number of 371 000 deaths is similar to the number for several of the top 10 causes of death in the USA in 2014, the year used by Newman-Toker *et al* in their analyses.⁸

CAN WE TRUST THESE NUMBERS?

People may question the set of data on which these conclusions are based as well as the validity of the assumptions and extrapolations to the total US population. So can we trust these numbers? The paper has a large number of supplementary files, providing detail on the methodology and statistical code used. A notable strength is that various sensitivity analyses were conducted to gauge the impact of possible overcounting and undercounting (appendix B) and several analyses to compare their estimates with those from other data sources or obtained using other methods (appendix C), suggesting these were roughly in the same range. For instance, they compare their estimates across care settings with previous estimates for inpatient settings^{9 10} to indicate that about 17% of the total serious misdiagnosis-related harms would occur in inpatient settings and 72 000 inpatient deaths. Other strengths include that the authors ensured every patient to be counted once to reduce possible overcounting (see below), that diagnostic error estimates were based on literature review of clinical studies and importantly, that final estimates of diagnostic error and harm rates were verified by experts on their face-validity.

The large amount of supplemental files may seem daunting, so we highlight here the most important aspects and limitations

to ensure proper interpretation of these estimates. The basic calculation is simple, where the authors multiply the serious misdiagnosis-related harm rate for a specific disease by the population incidence of that disease. This was done for 15 key diseases across 3 major disease categories (the 'Big Three'), previously found to account for three-quarters of serious harms in both malpractice claims and clinical studies (part 1 of the study).⁵

Considering the first element in the calculation, disease-specific misdiagnosis-related harm rates were taken from part 2 of the larger study, multiplying the diagnostic error rate by the harm rate per diagnostic error.⁷ Disease-specific diagnostic error rates were synthesised from clinical studies, but these studies mostly did not include misdiagnosis-related harm rates. Therefore, the authors used evidence from five studies reporting on the generic (disease-agnostic) misdiagnosis-related harms per diagnostic error, and used the average estimate of 30.8% (374 deaths or permanent disability from 1216 diagnostic errors) in all calculations, applying a disease-specific severity weight to take into account that a diagnostic error is likely more consequential for severe conditions such as aortic dissection than for less serious conditions. The severity weight was based on the disease-specific proportion of malpractice cases resulting in serious versus non-serious harms (see appendix A2). So, if for instance, the diagnostic error rate for a given disease was 10% and a severity weight of 2, then the misdiagnosis-related harm rate would be 6.16% per diagnostic error (ie, $2 \times 30.8\% = 61.6\%$, multiplied by the 10% with a diagnostic error). Because the 30.8% is used in all calculations, it is important to consider the five studies underlying that estimate, and the extent to which the included sample or methodology may have underestimated or overestimated the harm per diagnostic error.

Two studies were surveys—one a random sample (with 34% response rate) and the second a convenience sample—in which physicians were asked to recall a diagnostic error and the harm it had caused.^{11 12} It seems likely that participating physicians in these studies cared about this topic, and that recalled cases were ones with major harm. Such selection bias likely results in overestimating the true harm rate. One study relied on voluntary incident reports rather than being specifically conceived to study diagnostic errors, and unclear how the quality committee attributed harm to a diagnostic error.¹³ Voluntary incident reporting is known to capture only a fraction of events when compared with other methods such as record review, and may not reliably identify serious events, thereby underestimating the true harm rate (and likely why this study reported the lowest rate). Two other studies used record review with trigger tool methodology, with reviewer judgement to assess if the error contributed to or caused the harm.^{10 14} Using triggers selects

patients who were more likely to have experienced harm, thereby likely resulting in overestimated harm rates. Even though this oversampling was taken into account in reported diagnostic adverse event rates,¹⁰ Newman-Toker *et al* will likely not have had access to the weighting factor used to adjust for oversampling and therefore could only use the reported numbers. In addition, for at least some cases it would seem hard to judge whether the outcome for a patient would have been different had there not been a diagnostic error, for which we would need a control group. This is particularly relevant if these patients also experienced other adverse events that may or may not be related to the diagnostic error.¹⁰ As others have argued before, the fact that a preventable adverse event occurred close to the patient's death does not mean that the error is the cause of death, that is, the death could have been prevented.¹⁵ Therefore, it is more accurate to state that diagnostic errors have likely contributed to the harm and to refrain from more causal language.

The second element in the calculation is the population-based incidence of disease. Cancer registry data were used to estimate incidence of specific cancers. For vascular and infectious diseases, the authors counted discharge or in-hospital death diagnoses from national inpatient hospital stays and assumed that all patients diagnosed with these conditions in the outpatient setting (and those initially missed) would ultimately be hospitalised, to avoid double counting of patients, that is, first in the outpatient setting and then during hospitalisation. This seems a reasonable assumption for these 15 dangerous diseases. Furthermore, Newman-Toker *et al* conducted sensitivity analyses to gauge the impact of undercounting (by not including out-of-hospital deaths) and overcounting (by including patients with multiple hospitalizations, yet can die only once). These were found to have similar impact and therefore likely cancel each other out. However, the primary analysis includes both primary and secondary diagnoses and applies the above misdiagnosis-related harm rate to the total number, thereby implicitly assuming that harm resulting from the diagnostic error would be the same for diagnoses coded as primary or secondary diagnosis. We agree that a missed comorbidity for a patient admitted for another reason might also cause harm, but is likely less severe. We would argue that if a patient dies during admission or is discharged with permanent disability which is caused by a missed diagnosis, then this diagnosis would likely be listed as the primary discharge diagnosis or cause of death given its importance for the course of the admission. Furthermore, the number of secondary diagnosis codes is known to be high in the USA, likely influenced by financial incentives associated with coding.¹⁶ Therefore, we feel the sensitivity analysis only including the primary diagnoses is probably closer to the true number harmed, which reduced the overall estimates by 30% but still amounts to

diagnostic errors contributing to about half a million patients being seriously harmed and therefore does not change the overall message of the paper that this is an important problem that warrants action.

The final methodological issue relates to the combination of the two elements, that is, the extrapolation of study estimates to the total US population. We know from basic epidemiology that estimates involving associations in a specific study sample can only be generalised to the population from which the sample was taken. The authors acknowledge that some of the older studies may not relate to current practice, but that the diagnostic error rate seems stable or has even increased over time. However, we should also consider the distribution of diagnostic error across patient groups, where for instance a missed myocardial infarction—one of the 15 dangerous diseases examined by Newman-Toker *et al*—may occur more frequently in black patients,¹⁷ and the underlying social and structural determinants may increase the likelihood that it results in harm. In addition, out-of-hospital deaths were not included in the estimates by Newman Toker *et al* but may occur more frequently in some of these patient groups, meaning that harms from diagnostic errors could be underestimated for these patients. We should therefore only generalise to populations included in the studies on which the current estimates are based, and be cautious to assume they would be similar for patient groups under-represented in these studies. Other points related to the relatively small number of events of harm (n=374 across the five studies), meaning that uncertainty bounds around the misdiagnosis-related harm estimate will have been wide which is then extrapolated to the population. The authors acknowledge this and point to these currently being the best possible estimates that triangulate well with data from other sources and that experts deemed face valid. Finally, all calculations strictly speaking are only based on evidence for the ‘Big Three’ categories, and extrapolated to give a grand total of harm by dividing by 75.8%, assuming similar distributions and associations in the other quarter which may not be true. However, as above, the overall message when only including the studied diseases remains the same—diagnostic errors contribute to a large number of patients being harmed across care settings.

WHERE TO GO FROM HERE?

To better understand the harm caused by diagnostic error, we need more real-world measurements of specific types of diagnostic errors and harms, rather than additional studies based on assumptions and extrapolations. As with the overall adverse event rate,¹⁸ the causes, types and harms of diagnostic errors are extremely heterogeneous, and will require different interventions to remedy. These efforts should disentangle the harm from diagnostic error and the harm from other patient or health system factors.¹⁹

To do this, researchers should consider more robust study designs, for instance, employing propensity score matching techniques when using observational data to mimic randomisation in trials. This would give us better (average) estimates of harm in patients who experienced diagnostic errors versus those who did not. However, even then it would be hard to separate the effect of a diagnostic error initiating a cascade of other adverse events versus these clustering in the same (complex) patient, which will likely remain a matter of judgement. We might focus on diseases where most errors occur, identified by Newman-Toker *et al*. But within these diseases, we need to break things down by types and causes to really understand the mechanism and improve care.

Newman-Toker *et al* have shown that we have a very real and large problem that payers, regulators, accreditors, hospitals, healthcare organisations and physician groups should address. There has been substantial progress in the development of tools that enable and facilitate measurement of diagnostic errors, such as the ability to capture error reports from clinicians and patients and to use a growing array of e-resources based on the use of ‘trigger tools’.²⁰ Organisations can already start using those tools to identify where and why diagnostic errors occur for specific diseases. Finally, more research on interventions to prevent diagnostic errors is needed. Very few high-quality studies exist, most interventions were only tested in one site and many studies were small²¹ so we particularly need evidence whether such interventions have similar effects when scaled and replicated in other settings.

The past two decades have seen substantial progress in calling attention to the diagnostic error problem, and in clarifying how these arise from a wide range of interacting system-related and cognitive issues.³ We know how these arise from breakdowns in access, in communication and coordinating care, problems that get lost in the shuffle and the ever-growing array of cognitive ‘biases’ that characterise human thought and action. Each of these represents an opportunity to begin improving diagnosis and its outcomes. The aggregate harm that results from inaction should provide the call-to-action that is urgently needed to begin addressing the problem.

Twitter Eric J Thomas @EJThomas_safety

Contributors All authors contributed to conception of the paper, critically read and modified subsequent drafts and approved the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests PJM-vdM and EJT are editors at *BMJ Quality & Safety*. EJT receives research funding from the Agency for Healthcare Research and Quality to study diagnostic safety.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iDs

Perla J Marang-van de Mheen <http://orcid.org/0000-0003-1439-0989>

Eric J Thomas <http://orcid.org/0000-0002-3656-9181>

Mark L Graber <http://orcid.org/0000-0002-3600-0446>

REFERENCES

- Kohn LT, Corrigan JM, Donaldson MS. *To err is human: building a safer health system*. Washington, DC: National Academy Press, Institute of Medicine, 1999.
- Leape LL, Berwick D, Bates DW. Counting deaths due to medical error (authors reply). *JAMA* 2002;288:2405.
- National Academies of Sciences, Engineering, and Medicine. *Improving diagnosis in health care*. Washington, DC: The National Academies Press, 2015.
- Newman-Toker DE, Nassery N, Schaffer AC, et al. Burden of serious harms from diagnostic error in the USA. *BMJ Qual Saf*:bmjqs-2021
- 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.
- 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;8:127–8.
- 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.
- National vital statistics report. Deaths: final data for 2014. US Department of Health and Human Services; 2016. Available: https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_04.pdf
- 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.
- 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.
- Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;169:1881–7.
- Ely JW, Kaldjian LC, D'Alessandro DM. Diagnostic errors in primary care: lessons learned. *J Am Board Fam Med* 2012;25:87–97.
- Okafor N, Payne VL, Chathampally Y, et al. Using voluntary reports from physicians to learn from diagnostic errors in emergency medicine. *Emerg Med J* 2016;33:245–52.
- 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.
- Shojania KG, Dixon-Woods M. Estimating deaths due to medical error: the ongoing controversy and why it matters. *BMJ Qual Saf* 2017;26:423–8.
- O'Malley KJ, Cook KF, Price MD, et al. Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005;40(5 Pt 2):1620–39.
- Herasevich S, Soleimani J, Huang C, et al. Diagnostic error among vulnerable populations presenting to the emergency department with cardiovascular and cerebrovascular or neurological symptoms: a systematic review. *BMJ Qual Saf* 2023:bmjqs-2022-015038.
- Shojania KG, Marang-van de Mheen PJ. Identifying adverse events: reflections on an imperfect gold standard after 20 years of patient safety research. *BMJ Qual Saf* 2020;29:265–70.
- Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in Healthcare: the safer DX framework. *BMJ Qual Saf* 2015;24:103–10.
- Bradford A, Singh H. Measure DX; A resource to identify, analyze, and learn from diagnostic safety events. Agency for Health Care Research and Quality (AHRQ); 2022. Available: <https://www.ahrq.gov/patient-safety/settings/multiple/measure-dx.html>
- Dave N, Bui S, Morgan C, et al. Interventions targeted at reducing diagnostic error: systematic review. *BMJ Qual Saf* 2022;31:297–307.