The Diagnostic Performance Feedback "Calibration Gap": Why Clinical Experience Alone Is Not Enough to Prevent Serious Diagnostic Errors

Rodney Omron, MD, MPH, Susrutha Kotwal, MD, Brian T. Garibaldi, MD, and David E. Newman-Toker, MD, PhD

Medical diagnostic errors can be thought of as the bottom of the iceberg of patient safety—a hidden yet vast source of morbidity and mortality.[1] According to the U.S. National Academy of Medicine, diagnostic errors represent a major public health problem likely to affect each of us in our lifetime.[2] Diagnostic errors contribute to approximately 10% of deaths and 6% to 17% of hospital adverse events and are the leading cause of medical malpractice claims.[3] Although typically multifactorial, the majority of diagnostic errors can be traced back to failures in bedside examination skills and clinical reasoning;[4] knowledge and skill gaps appear to play important roles that have been underestimated in the context of an overemphasis on cognitive bias as a cause.[5]

A critical unanswered question for educational strategies to improve diagnosis is how diagnostic errors could remain so common, even with clinical presentations seen daily in clinical practice. In theory, accumulated clinical experience gained over time should be an "antidote" that gradually eliminates misdiagnosis, but experimental studies suggest that more years of experience does not necessarily confer greater diagnostic accuracy.[6] Part of the problem is that feedback is essential for improved diagnostic performance, but is often lacking.[7] For example, it has been shown that some short-term deaths after discharge from the emergency department (ED) likely reflect missed diagnoses of life-threatening illnesses.[8] Unfortunately, such feedback rarely returns to individual ED clinicians, which prevents "recalibration" (i.e., adjusting one's own mental models for diagnosis based on real-world accuracy in prior similar cases) that would otherwise improve diagnostic performance.[9] In this article we argue that, absent systematic feedback, even years of sustained clinical practice may not produce the necessary experiential learning to prevent critical diagnostic errors, particularly for high-risk, low-frequency conditions.

As a representative case, take the known public health problem of missed stroke in patients presenting with acute dizziness to the ED, where 45,000 to 75,000 strokes are missed at first contact each year[10] and an estimated 10,000 to 25,000 serious preventable harms result from missed opportunities for early treatment.[10] Dizziness and vertigo are common problems, accounting for approximately 3% of all ED

From the Johns Hopkins University School of Medicine (RO, SK, BTG, DENT), Baltimore, MD.
Received July 4, 2018; accepted July 28, 2018.
Dr. Newman-Toker's effort was supported by a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD, #U01 DC013778) and the Armstrong Institute Center for Diagnostic Excellence.
DENT serves as a paid consultant, reviewing medicolegal cases for both plaintiff and defense firms related to misdiagnosis of neurologic conditions, including dizziness and stroke. He has conducted government and foundation funded research related to diagnostic error, dizziness, and stroke. He has been loaned research equipment related to diagnosis of dizziness and stroke by two commercial companies (GN Otometrics and Interacoustics) and Johns Hopkins has licensed related diagnostic decision-support technology to GN Otometrics for which DENT receives royalties.
Author contributions: DENT created the study concept; RO and SK drafted the manuscript; and BTG, SK, and DENT helped RO with the critical revision of the manuscript.
Supervising Editor: Susan Promes, MD.
Address for correspondence and reprints: Rodney Omron, MD, MPH; e-mail: romron1@jhmi.edu.
AEM EDUCATION AND TRAINING 2018;2:339–342

Table 1
Calibration Gap for Missed Stroke in Dizziness for an Average Full-time Clinical ED Physician

Parameter	No. per year	No. per decade
ED visits	60,000	600,000
Dizziness visits (~3%)	1,800	18,000
Dizziness patients while 'on shift' (~16%–18%)	300	3,000
Dizziness while on shift due to stroke (~3%–5%)	12	120
Obvious strokes (e.g., with neurologic signs) recognized (~60%*)	7	70
Subtle strokes (e.g., isolated vertigo with only eye/gait signs) missed (~40%*)	5	50
Untreated missed strokes returning with major stroke < 90 days (~15%–20%)	1	10
Missed strokes returning with major stroke to the same hospital (~60%)	< 1	6
Missed strokes returning with major stroke while "on shift" (~16%–18%)	~0	1
Total missed strokes (actual count)	5	50
Total missed strokes w/direct feedback for calibration (perceived count)	~0	1

*The precise frequency of ED stroke misdiagnosis in dizziness is unknown, but best estimates suggest it is ~40%[11]

visits, but only approximately 3% to 5% of these are due to stroke.[10] A typical full-time clinical ED physician at a medium-sized hospital ED (~60,000 visits per year) will see about 3,000 patients with dizziness each decade, 120 due to stroke (Table 1, Figure 1). Roughly 80% will have isolated vestibular symptoms without general neurologic signs[12] and a large fraction of these will be misdiagnosed with inner ear problems (or some other form of presumptively "benign" dizziness) and discharged. Only approximately 15% to 20% will go on to have major stroke in the days and weeks after minor stroke or transient ischemic attack[13] and roughly 40% of these will return to a different ED in the region[14] (assuming that they do not die before making it to the hospital). So, absent systematic follow-up, an ED physician with a 40% miss rate[11] might find out about only approximately one dizzy stroke patient per decade they personally called "benign," despite having missed 50 such strokes. Miscalibration occurs because that physician's perceived experience is of having recognized 70 (obvious) strokes and only missing one (not-so-obvious) stroke, giving an internal mental-model estimate of 99% sensitivity, despite an actual sensitivity of 60%. Varying the estimated diagnostic accuracy from 20% to 80% for dizzy strokes does not change the bottom line—everyone overestimates their accuracy when they do not get feedback about the cases they miss.[15] This leads to a massive "calibration gap" that likely keeps diagnostic performance stagnant over time (and to a general misconception on the part of clinicians that such errors are very rare events).

This calibration gap could obviously be narrowed by "closing the loop" on diagnostic performance

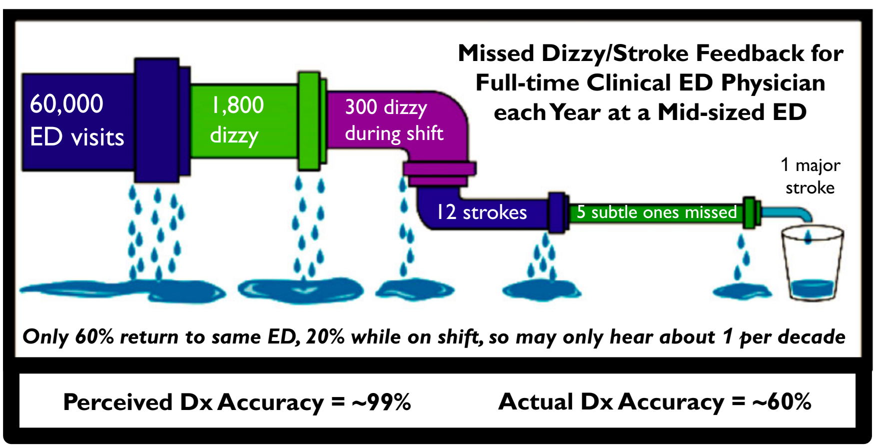

Figure 1. Leaky pipeline of feedback on diagnostic performance. There is an enormous gap between actual diagnostic performance and perceived diagnostic performance, because of loss to follow-up. Initial artwork modified from http://www.techvision21.com/the-bachelors-to-ph-d-pipeline-is-not-leaking-women-and-underrepresented-minorities/). Dx = diagnosis.

through systematic, prompt, personalized feedback on diagnostic process, accuracy, error, and harms for all clinical cases. As yet, however, no effective mechanism exists for providing such routine, operational feedback at the individual provider level.[16] Furthermore, the relative infrequency of cases due to dangerous disorders such as stroke would likely produce very slow learning, over years or decades. While live patient experiences with full, immediate follow-up on diagnostic performance (both process and outcomes) are probably the most potent for achieving learning and expertise,[17] the overall low prevalence of dangerous disorders makes it nearly impossible (even with perfect feedback) for such expertise to develop efficiently.

An alternative would be to leverage novel simulation-based educational approaches in a blended learning format to achieve expertise through deliberate (topic-focused) and mixed (epidemiologically valid) practice.[18,19] The theory of deliberate practice states that improved performance emerges when one engages with the content, using deliberate effort, followed by feedback on diagnostic performance.[20] Simulation-based medical education with deliberate practice has been demonstrated to be superior to traditional clinical education.[21] Case-based computer simulations (often known as "virtual patients"[22]) are increasingly being used to enhance clinical education, and libraries of such patients could be used as a resource to support both deliberate and mixed practice with a goal of improving diagnostic expertise.[17] Many virtual patients are "fake" patients developed as educationally idealized versions of specific disorders, but useful learning is enhanced when cases and associated diagnostic decisions are closer to what is encountered in clinical practice,[23] so virtual patient libraries should be as "real" as possible. Although difficult to create, such libraries would be readily scalable after initial construction.

One mechanism to develop "real" virtual patient libraries would be to use digital sources of real-world data, such as those systematically obtained as part of standardized clinical examinations and test batteries (e.g., device-enhanced "tele-dizzy" consultations in the ED[24]) or diagnostic clinical trials (e.g., Acute Video-oculography for Vertigo in Emergency Rooms for Rapid Triage [AVERT], NCT02483429). Structured clinical histories and sensor-derived physical examinations (e.g., eye movement recordings using video-oculography[25]) could be deployed as virtual patients in existing virtual case platforms (e.g., the Virtual Interactive Case System[26] or i-Human[27]). Furthermore, "digital" partial task trainers for use on mobile devices (e.g., aVOR[28]) could be linked to these virtual patient cases to develop the psychomotor aspects of fundamental physical examination skills in parallel to critical cognitive skills of interpreting and integrating findings.

While this may sound futuristic, it may not be far off for some clinical presentations. The above-mentioned clinical trial of video-oculography to diagnose acute dizziness and vertigo in the ED (AVERT) is now collecting structured history and physical examination findings from patients who present to the ED with symptom of dizziness vertigo and have examination findings suggestive of inner ear or brain dysfunction. Trial materials include actual patient videos of eye movements and neurologic examinations, as well as fully vetted final diagnoses adjudicated by a multidisciplinary team using strong criterion standard assessments. The authors have already begun to create a library of virtual patient training cases that will then be tested first for their impact on simulated performance using similar cases and, eventually, real-world diagnostic performance.

Similar approaches could be taken for other presentations known to be associated with high rates of diagnostic error and harm (e.g., chest pain/pulmonary embolus, fever/sepsis, back pain/spinal epidural abscess). Because most of the misdiagnosis-related harms in the ED are likely due to a relatively small number of high-risk clinical presentations and diseases (mostly vascular events and bacterial infections[29]), simulation-based training for such presentations is probably realistic. Developing case materials may take time and effort, but demonstration projects such as the one we propose for dizziness and stroke could guide the way.

Awareness of diagnostic error frequency, preventability of harms, and the "calibration gap" problem is an essential first step in moving clinicians from competence toward excellence in diagnosis, a goal valued by the entire health care team.[30] Relying on "accumulated years of clinical experience" is not currently meeting this need, and novel educational approaches are likely needed to achieve diagnostic excellence in everyday clinical practice.

References

1. Newman-Toker DE, Tucker L; SIDM Policy Committee. Roadmap for Research to Improve Diagnosis, Part 1: Converting National Academy of Medicine Recommendations into Policy Action. Evanston (IL): Society to Improve Diagnosis in Medicine, 2018.

2. Improving Diagnosis in Healthcare. National Academy of Medicine; 2015. Available at: http://www.nationalacademies.org/hmd/Reports/2015/Improving-Diagnosis-in-Healthcare.aspx. Accessed Jun 5, 2018.
3. National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press, 2015.
4. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med 2009;1691881–7.
5. Dhaliwal G. Premature closure? Not so fast. BMJ Qual Saf 2017;26:87–89.
6. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. JAMA Intern Med 2013;173:1952–8.
7. Schiff GD. Minimizing diagnostic error: the importance of follow-up and feedback. Am J Med 2008;121(5 Suppl):S38–42.
8. Obermeyer Z, Cohn B, Wilson M, Jena AB, Cutler DM. Early death after discharge from emergency departments: analysis of national US insurance claims data. BMJ 2017;356:j239.
9. Croskerry P. The feedback sanction. Acad Emerg Med 2000;7:1232–8.
10. Newman-Toker DE. Missed stroke in acute vertigo and dizziness: it is time for action, not debate. Ann Neurol 2016;79:27–31.
11. Tarnutzer AA, Lee SH, Robinson KA, Wang Z, Edlow JA, Newman-Toker DE. ED misdiagnosis of cerebrovascular events in the era of modern neuroimaging: a meta-analysis. Neurology 2017;88:1468–77.
12. Tarnutzer AA, Berkowitz AL, Robinson KA, Hsieh YH, Newman-Toker DE. Does my dizzy patient have a stroke? A systematic review of bedside diagnosis in acute vestibular syndrome. CMAJ 2011;183:E571–92.
13. Rothwell PM, Buchan A, Johnston SC. Recent advances in management of transient ischaemic attacks and minor ischaemic strokes. Lancet Neurol 2006;5:323–31.
14. Vermeulen MJ, Schull MJ. Missed diagnosis of subarachnoid hemorrhage in the emergency department. Stroke 2007;38:1216–21.
15. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008;121(5 Suppl):S2–23.
16. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. BMJ Qual Saf 2018;27:557–66.
17. Sherbino J, Norman GR. Reframing diagnostic error: maybe it's content, and not process, that leads to error. Acad Emerg Med 2014;21:931–3.
18. Duvivier RJ, van Dalen J, Muijtjens AM, Moulaert VR, van der Vleuten CP, Scherpbier AJ. The role of deliberate practice in the acquisition of clinical skills. BMC Med Educ 2011;11:101.
19. Hatala RM, Brooks LR, Norman GR. Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. Adv Health Sci Educ Theory Pract 2003;8:17–26.
20. Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. The Cambridge Handbook of Expertise and Expert Performance. New York: Cambridge University Press, 2006.
21. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med 2011;86:706–11.
22. Kononowicz AA, Zary N, Edelbring S, Corral J, Hege I. Virtual patients–what are we talking about? A framework to classify the meanings of the term in healthcare education. BMC Med Educ 2015;15:11.
23. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ 2005;39:98–106. Review. Erratum in: Med Educ 2005 Jul;39:753.
24. Gold D, Tourkevich R, Brune A, et al. A Novel Tele-Dizzy Consultation Program in the Emergency Department Using Portable Video-oculography. 30th International Barany Society meeting, Uppsala, Sweden, 2018.
25. Newman-Toker DE, Curthoys IS, Halmagyi GM. Diagnosing stroke in acute vertigo: the HINTS family of eye movement tests and the future of the "eye ECG". Semin Neurol 2015;35:506–21.
26. Virtual Interactive Case System. c2008-2013. Available at: http://www.pie.med.utoronto.ca/VIC/index.htm. Accessed Jul 3, 2018.
27. i-Human Patients. c2008-2018. Available at: http://www.i-human.com/. Accessed Jul 3, 2018.
28. AVOR – Vestibular Education APP Has Reached 10,000 Downloads. c2002-2017. Available at: http://sydney.edu.au/science/psychology/news/news20130306-1. Accessed Jul 3, 2018.
29. Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. Ann Emerg Med 2007;49:196–205.
30. Mote PC, Solomon BS, Wright SM, Crocetti M. Clinical excellence in pediatrics. Clin Pediatr 2014;53:879–84.