

The data of diagnostic error: big, large and small

Gurpreet Dhaliwal,^{1,2} Kaveh G Shojania³

¹Department of Medicine, University of California, San Francisco, San Francisco, California, USA

²Medical Service, San Francisco VA Medical Center, San Francisco, California, USA

³Department of Medicine and Centre for Quality Improvement and Patient Safety, University of Toronto, Toronto, Ontario, Canada

Correspondence to

Dr. Gurpreet Dhaliwal, Department of Medicine, University of California, San Francisco CA 94121, USA; gurpreet.dhaliwal@ucsf.edu

Accepted 16 February 2018

Diagnostic error research has mostly focused on methods to detect, characterise and analyse lapses in the diagnostic process by using incident reports, malpractice claims, autopsies and electronic trigger tools. The associated literature shows how frequently important diagnostic errors occur¹ and examines cognitive² and system-based³ causes of these errors. Relatively absent from this portfolio of research have been large-scale approaches for measuring institutional diagnostic performance, either for benchmarking purposes or for driving improvement efforts.

In this issue of *BMJQS*, Liberman and Newman-Toker introduce Symptom–Disease Pair Analysis of Diagnostic Error (SPADE) as a new approach to identify diagnostic errors by analysing large patient data sets (tens of thousands of patient encounters housed in electronic medical records or administrative databases).⁴ The SPADE methodology starts with a symptom that is misdiagnosed at an appreciable rate such as chest pain or dizziness. It then looks for instances within the data set where a patient with that symptom has two coded encounters in a short time frame. Misdiagnosis-related harm is inferred when there is a *prespecified change in diagnosis over time*. An example is a patient with acute dizziness (the symptom) who is discharged with a diagnosis of positional vertigo at an initial emergency department (ED) encounter and 1 week later returns to the ED and is diagnosed with an acute stroke (the disease).

The Symptom–Disease Pair at the core of the SPADE methodology refers to the presenting symptom (eg, dizziness) and the correct, but initially overlooked, diagnosis (eg, stroke). The validity of any symptom–diagnosis pair as a marker of diagnostic error is established using two approaches. A ‘look

forward’ approach begins by identifying all patients who presented with dizziness and were discharged with a diagnosis of benign positional vertigo. It then looks at how often these patients present again within 30 days and receive a stroke diagnosis. A ‘look back’ approach establishes the proportion of patients admitted with stroke who presented within the preceding 30 days with dizziness. The SPADE approach assesses the frequency of this symptom–disease dyad across a large data set and counts each instance as an episode of misdiagnosis-related harm.

SPADE is an advance in diagnostic error measurement. It focuses on organisation-wide, disease-specific, misdiagnosis-related harm. It provides an estimate of diagnostic error frequency without human adjudication of flagged cases. And it offers a dynamic measure of diagnostic performance that can be followed over time. A companion short report and video in this issue of *BMJQS* illustrate how the SPADE approach can be used to populate an interactive diagnostic performance dashboard.⁵

SPADE, however, will capture only a subset of diseases, and within those conditions, only a subset of the diagnostic error burden. The SPADE methodology is best suited to detect diagnostic harm associated with symptoms where there is an elevated risk of acute worsening in the short term (eg, chest pain). There also must be a strong bidirectional statistical link between one symptom and one disease for SPADE to infer diagnostic error with confidence. Change in diagnosis must take place over two *discrete* episodes to be detected by SPADE. If a patient is misdiagnosed and later correctly diagnosed during a *single* episode of care (eg, within a hospitalisation), the change in diagnosis will not be detected. For instance, a patient admitted with syncope attributed to hypovolaemia and



► <http://dx.doi.org/10.1136/bmjqs-2017-007032>



To cite: Dhaliwal G, Shojania KG. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2018-007917

diagnosed days later during the same hospitalisation with ventricular tachycardia will not generate a signal of delayed diagnosis. These concerns, however, should be tempered by the understanding that organisations do not need a perfect account of diagnostic error in order to initiate improvement efforts. A general sense of the magnitude can spark change.

A recurrent theme in patient safety is that no single method provides the complete picture of the problem.⁶ Whether we are trying to identify cases involving diagnostic errors, generating strategies for avoiding diagnostic errors or measuring diagnostic performance, several complementary approaches are needed.⁷ For the part of the puzzle that relates to acute conditions with tight linkages between a single symptom and single morbid diagnosis, the SPADE methodology is a promising tactic.

LARGE DATA

The authors affiliate SPADE with ‘big data’, a term with an increasingly uncertain meaning on account of its ubiquity and hype. SPADE certainly depends on large data sets, but its data sources and analytical methods more closely resemble the size and types used in traditional health services research. Big data more distinctly refers to data sets that are constructed from multiple structured and unstructured sources and are often so voluminous and complex that traditional data processing software are inadequate to analyse them; oftentimes advanced computational methods, including artificial intelligence and machine learning, are necessary to gain insights from the information. In practice, the demarcation between traditional large data set analysis and big data analytics is fuzzy. Although SPADE may not count as ‘big data’ in the strictest sense, it nonetheless represents a foundational step that illustrates how diagnostic error investigators can leverage large data sets to identify important targets for improvements in diagnostic performance.

A promising starting point will be for large organisations to pilot a SPADE analysis with disease-specific diagnostic errors that have evidence-based solutions. For instance, if this approach unearths a rising number of patients with misdiagnosed spinal epidural abscess, an institution could implement a protocol incorporating risk factor assessment followed by testing for erythrocyte sedimentation rate and C-reactive protein, which has decreased diagnostic delays for this condition.⁸ Yet, even if SPADE creates a signal (and the system institutes a new protocol), there remains a risk that many clinicians will react without modifying their practice. What can help motivate that change?

SMALL DATA

Enter small data. ‘Small data’ is information that is comprehensible without analytics and comes in a volume and format that makes it manageable and informative.^{9 10} The most accessible version of small

data is a story.¹¹ And the most compelling stories for clinicians concern their own patients.

But the only way for clinicians to *learn* from their patient stories is to learn how each story *ends*.¹² Countless patient encounters end with provisional or unknown diagnoses, where definitive diagnostic confirmation (or refutation) occurs days, weeks or months later as testing, treatment and natural history play out.¹³ Clinicians who wish to optimise their diagnostic judgements must establish patient tracking systems where they learn how the story ends—and how they can do better the next time they encounter the same diagnostic problem.¹⁴ This small data approach can help catalyse the change that big data output brings to our attention.

Suppose SPADE surfaces increasing rates of delayed diagnoses of spinal epidural abscess in elderly patients. As clinicians, we would certainly take an interest in this system-wide signal. But what specifically do we do the next time we are confronted with an octogenarian with back pain? To generate a plan for individual improvement, a clinician would need to analyse their own cases to see where they deviated from best practices and discern other shortcomings in their diagnostic approach.¹⁵ Such examination of small data could lead to insights about insufficient examination techniques, underestimation of the frequency of MRI misinterpretation or overestimation of the utility of fever or leukocytosis.

This small data complement to big data approaches will be particularly important with commonly misdiagnosed conditions like pneumonia and cellulitis¹⁶ where the causes of misdiagnosis are heterogeneous and complex and where few prepackaged solutions exist. Big data can surface the problem, but small data will provide the insights and motivation to do something about it.

CONCLUSION

Current research methods are insufficient to understand the magnitude and causes of diagnostic error. SPADE is not yet ready for high-stakes external benchmarking, but it can be piloted within large organisations to see if reliable internal metrics that drive improvements in diagnostic safety can be achieved. Studying diagnostic error is incredibly complex, and it will take time to develop methods. The SPADE approach is an important step forward.

Organisations should welcome large data sets into the portfolio of approaches to detect and measure diagnostic error. But they cannot lose sight of this persistent problem in efforts to improve quality: that data alone do not change practice. Patient stories have often found a place in efforts to motivate improvement efforts,^{17 18} as they are both emotionally and intellectually engaging.¹⁹ Clinicians who set up their own tracking systems quickly learn that the most powerful stories are ones where their own patients are

the protagonists and that patient-specific rather than population-wide feedback is the strongest motivator for improvement.²⁰

The curation and sharing of these stories should remain a priority even with the rise of big data. Technology has made amazing progress in recent decades, but the human brain has not changed one bit. The story, not the statistic, remains the brain's preferred unit of learning—and the most powerful tool of persuasion.

Contributor GD and KGS contributed to the conception of the paper; they critically read and modified subsequent drafts and approved the final version. KGS is an editor at BMJ Quality & Safety.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests GD reports receiving honoraria from ISMIE Mutual Insurance Company and Physicians' Reciprocal Insurers.

Provenance and peer review Commissioned; internally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf* 2014;23:727–31.
- Graber ML, Kissam S, Payne VL, *et al.* Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012;21:535–57.
- Singh H, Graber ML, Kissam SM, *et al.* System-related interventions to reduce diagnostic errors: a narrative review. *BMJ Qual Saf* 2012;21:160–70.
- 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. doi: 10.1136/bmjqs-2017-007032. [Epub ahead of print 22 Sep 2018].
- Mane *et al.* Diagnostic Performance Dashboards—Tracking Diagnostic Errors using Big Data. *BMJ Qual Saf*. bmjqs-2018-007945.
- Levtzion-Korach O, Frankel A, Alcalai H, *et al.* Integrating incident data from five reporting systems to assess patient safety: making sense of the elephant. *Jt Comm J Qual Patient Saf* 2010;36:402–AP18.
- Shojania KG. The elephant of patient safety: what you see depends on how you look. *Jt Comm J Qual Patient Saf* 2010;36:399–AP3.
- Davis DP, Salazar A, Chan TC, *et al.* Prospective evaluation of a clinical decision guideline to diagnose spinal epidural abscess in patients who present to the emergency department with spine pain. *J Neurosurg Spine* 2011;14:765–70.
- 10 Reasons 2014 will be the Year of Small Data. <http://www.zdnet.com/article/10-reasons-2014-will-be-the-year-of-small-data/> (accessed 3 Feb 2018).
- Sacristán JA, Dilla T. No big data without small data: learning health care systems begin and end with the individual patient. *J Eval Clin Pract* 2015;21:1014–7.
- Aronson L. A piece of my mind. Story as Evidence, Evidence as Story. *JAMA* 2015;314:125–6.
- Bowen JL, Ilgen JS, Irby DM, *et al.* "You Have to Know the End of the Story": Motivations to Follow Up After Transitions of Clinical Responsibility. *Acad Med* 2017;92:S48–S54.
- Schiff GD. Minimizing diagnostic error: the importance of follow-up and feedback. *Am J Med* 2008;121:S38–S42.
- Dhaliwal G. Annals for Hospitalists Inpatient Notes - Diagnostic Excellence Starts With an Incessant Watch. *Ann Intern Med* 2017;167:HO2–HO3.
- Bhise V, Meyer AND, Singh H, *et al.* Errors in Diagnosis of Spinal Epidural Abscesses in the Era of Electronic Health Records. *Am J Med* 2017;130:975–81.
- Singh H, Giardina TD, Meyer AN, *et al.* Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013;173:418–25.
- Wachter RM, Shojania KG, Markowitz AJ, *et al.* Quality grand rounds: the case for patient safety. *Ann Intern Med* 2006;145:629–30.
- Stang AS, Wong BM. Patients teaching patient safety: the challenge of turning negative patient experiences into positive learning opportunities. *BMJ Qual Saf* 2015;24:4–6.
- Cox K. Stories as case knowledge: case knowledge as stories. *Med Educ* 2001;35:862–6.
- Mitchell E, Sullivan F, Grimshaw JM, *et al.* Improving management of hypertension in general practice: a randomised controlled trial of feedback derived from electronic patient data. *Br J Gen Pract* 2005;55:94–101.