

Finding and fixing diagnosis errors: can triggers help?

Gordon D Schiff

Imagine conferring with your clinician colleagues and being handed a plateful of all of your missed and delayed diagnoses. But, imagine further that, rather than a nightmare of ghosts returning to haunt you in the form of malpractice claims, sanctions by regulatory boards, insurers pouncing on needless expenditures or hordes (yes, there would be large numbers) of angry finger-pointing patients and families, the experience would instead bring a dream of supportive feedback and learning. Imagine the ways such an idealised non-threatening consultation and conference might be designed to minimise defensiveness and maximise introspection, learn lessons, and rethink habits and standard practices. Rather than prompting incredulous exclamations of “you missed that?!” or “what were you thinking?!” the process would generate engagement and scrutiny of office and hospital workflow and diagnostic testing practices, realistically grappling with time and cost trade-offs, pressures, uncertainties and diagnostic challenges that practicing clinicians face every day. In short, your plateful of missed diagnoses would initiate a process that combines the best elements of a fun and informative morbidity and mortality conference, an expert second opinion from a generous colleague and productive quality improvement consultation.

Correspondence to Dr Gordon D Schiff, Associate Director, Center for Patient Safety Research and Practice Division of General Internal Medicine, Brigham and Women's Hospital Associate Professor of Medicine Harvard Medical School 1620 Tremont St. 3rd Fl Boston, MA 02120; USA; gschiff@partners.org

I suspect the average clinician could care less about diagnostic ‘triggers’ or a new study to increase their positive predictive value. However, no professional could fail to see the appeal of the ultimate form of continuing medical education imagined above—learning practical lessons from one’s own cases and discussing with trusted colleagues the ways care could be improved. How to get there from here poses a fundamental challenge, one that the study by Singh *et al* on diagnosis error ‘triggers’ in this issue of *BMJ Quality and Safety* attempts to address.¹

The authors, a research team based at the DeBakey Veterans Affairs Medical Center and Baylor College of Medicine in Houston, have continued to refine their approaches to reliably and efficiently screen for diagnostic error cases.^{1 2} Using the comprehensive longitudinal electronic medical record in the US Veterans Health Administration hospitals (and now another electronic medical record in a second health system), they took advantage of two powerful features of electronic records that will increasingly change the paradigm of quality improvement and health services research: (a) the ability to perform simple or even complex queries to select samples of patients meeting specified criteria and (b) the ability to efficiently retrieve and review electronic patient records identified by the screening criteria. Thus, the next decade of ‘chart review’ (which, as the authors point out, is essential for evaluating potential diagnosis error cases), with this quantitatively eased burden of case selection and review, holds the

promise of facilitating qualitative leaps forward in what we can learn from the details of clinical care processes and outcomes.

As the authors and the literature repeatedly point out, diagnostic errors are important but understudied, largely because of difficulties in defining and detecting such errors.^{3 4}

Can generic electronic screens cast an effective and efficient net to pick out/up charts to review for errors in diagnosis with a sufficiently high likelihood of errors (ie, positive predictive value) to make the manual reviews worth the effort? In attempting to ease the burden of detecting diagnostic errors, the authors (and others) have sought to develop methods to more efficiently sift through the tens of thousands of encounters they wanted to screen for diagnostic errors. In the present study,¹ they developed two relatively simple screens—admission to an acute care hospital in the 14 days following an index primary care visit, or a second broader screen that identified patients who presented to an emergency department, sought urgent care or had an unscheduled subsequent primary care visit, also in the 14 days following an index visit.

Their rationale, that an unexpected hospitalisation, emergency department visit or urgent/unscheduled clinic visit could result from a missed diagnosis during the initial visit, seems sound, particularly for acute problems. Unfortunately, it overlooks considerable numbers of other situations where diagnosis fails. In fact, it seems unlikely to pick up many of the cases that represent the leading causes of malpractice claims in primary care—missed and delayed diagnoses of cancer. Here the diagnostic failures unfold over a period of weeks and months, rather than hours or days, and, except in relatively rare circumstances (eg, progression of an undiagnosed colon cancer to the point of causing bowel

obstruction), would not typically lead to a positive trigger such as hospitalisation within the subsequent 2 weeks. The authors acknowledge this limitation and have, in other studies,^{5 6} sought complementary screens to detect missed cancer diagnoses. But this shortcoming illustrates the fact that no one screen fits all when it comes to detection of diagnostic errors.

More subtle, but sobering, is the evidence from this study's data that their screen would fail to detect more than 90% of the missed diagnoses in their cohort. Using the trigger increased the yield from roughly 2% of unselected ('control') cases, to 20% in the trigger-positive sample (meeting the 14-day hospitalisation screening criteria). However, applying this 2% rate to their cohort of screen-negative patients means that while they picked up a total of 177 error cases in the 'screen positive' patients (using both sets of trigger screens), there would have been 1710 additional cases involving missed diagnoses—nearly 10 times as many error cases in the larger cohort overlooked with a negative screen. Depending on the purpose of the error screening, this may or may not be acceptable. Certainly, if I wanted to hear about all of my cases of diagnostic error, I would find my plate 90% empty. On the other hand, if an institution seeks a quick sample of cases to illustrate problems with diagnosis, using the selective triggers to target a manageable number of charts to review might prove quite helpful.

Another important question this sampling strategy raises is the representativeness of the cases found via the screen. In other words, are the 177 cases identified by the triggers and subsequent chart reviews, fairly similar to or systematically different from the 1710 overlooked cases? As mentioned above, to the extent that they overlook missed diagnoses more related to chronic or subacute

diseases, the triggers may miss the identification of important system problems related to diagnostic processes. As long as the institution recognises the characteristics of different possible triggers, it could probably develop an efficient, trigger-based strategy for identifying substantially delayed diagnosis of cancers and other subacute illnesses, as opposed to more acute problems.

ARE TRIGGERS THE ANSWER?

'Triggers' have become an increasingly used buzzword in adverse event detection and measurement.⁷⁻⁹ The Institute for Healthcare Improvement has developed and promoted a 'global trigger tool' as an instrument 'to surface harm' and for measuring rates of injury from medical care over time.^{10 11} In 2008, the Agency for Healthcare Research and Quality convened an expert panel meeting on triggers and targeted injury detection systems to review the trigger literature and discuss challenges in implementing targeted injury detection systems.⁷ And the Singh study uses the word 'trigger' 67 times (excluding the abstract and references) to reference their screening criteria.¹

However, more than a century ago, Sir William Osler, the great teacher and diagnostician, promulgated a different use of the word 'trigger'.¹² He urged his students to use 'triggers' from the history and physical to create a list of possible diagnoses and then narrow the differential diagnosis based on their knowledge of anatomy and physiology. How effectively do today's clinicians, still largely relying on unaided human memory, trigger the correct diagnosis? A study that my colleagues and I recently published¹³ suggests that such 'triggering' represents a major weak spot and contributor to diagnosis error. In fact, failure or delay in *considering* the diagnosis constituted the leading factor, found

in 110 of 583 cases in a series of self-reported diagnosis errors, far exceeding the second most frequent (and also related to failure to consider) failure mode—failure/delay in ordering needed tests (seen in 63 cases).

Dr Lucian Leape and I have recently proposed six components, ideally aided by computerised decision support in some form, as a model for effectively triggering the thought processes and actions essential for reliable diagnosis.¹⁴ These elements for fail-safe diagnostic assessment of a given symptom or problem include: (1) key data elements to be collected (ideally automatically via computer assisted questionnaires); (2) 'don't miss' diagnoses—critical diagnoses that one must consider given their seriousness and need for urgent treatment (eg, aortic dissection in patients presenting with acute chest pain); (3) "red flag" symptoms and signs suggesting a potential 'don't miss diagnosis' (eg, back pain that wakens a patient at night); (4) potential drug causes (because clinicians frequently overlooked medication-related causes of patients' symptoms); (5) required referral(s) (ie, when to refer for more expert evaluation or diagnostic procedures); and (6) patient follow-up instructions and plans (so the patient knows what to watch for and when to follow-up if they do not feel better).

Designing systems (including information technologies, but other structures and processes of care as well) to hardwire these elements into clinical workflow and diagnosis decision support holds the promise of substantially reducing delayed or missed diagnoses due to the problem of overlooking relevant diagnostic considerations. While there are many other challenges in the diagnostic process (box 1), such as test results lost to follow-up (which many consider a lower hanging fruit candidate for improving diagnosis¹⁵),

Box 1 Reliable diagnosis: pitfalls and challenges**Challenging disease presentation**

Atypical presentation
 Non-specific symptoms and signs
 Unfamiliar/outside specialty
 Findings masking/mimicking another diagnosis
 Red herring misleading findings
 Rapidly progressive course
 Slowly evolving blunting onset perception
 Deceptively benign course

Patient factors

Language/communication barriers
 Signal:noise—patients with multiple other symptoms or diagnoses
 Failure to share data (to be forthcoming with symptoms or their severity)
 Failure to follow-up

Testing challenges

Test not available due geography, access, cost
 Logistical issues in scheduling, performing
 False positive/negative test limitations
 Performance/interpretation failures
 Equivocal results/interpretation
 Test follow-up issues (eg, tracking pending results)

Stressors

Time constraints for clinicians and patients
 Discontinuities of care
 Fragmentation of care
 Memory reliance/challenges

Broader challenges

Recognition of acuity/severity
 Diagnosis of complications
 Recognition of failure to respond to therapy
 Diagnosis of underlying etiologic cause
 Recognising misdiagnosis occurrence

The online supplementary appendix 1 presents the above contributing factors in a grid format along with commonly missed diagnoses.

wrong' represents a fundamental safety requirement.¹⁷ For high reliability diagnosis, such worrying requires that clinicians have a heightened 'situational awareness' of where the Swiss cheese holes¹⁸ lie in diagnosis. We have drafted a matrix (online supplementary appendix 1) to illustrate some of the more frequent, critical, problem-prone diagnoses we observed during a series of diagnosis, morbidity and mortality discussions as well as from several hundred self-reported errors.¹³ There is a need to recognise the varying mix of factors at play in different diagnoses and situations. This matrix illustrates how the vulnerabilities listed in **box 1** that populate the columns of the grid can conspire to make diagnosis difficult. Such a failure-mode vulnerability checklist can be used both retrospectively—to analyse potential error cases—and prospectively—to heighten awareness of pitfalls in approaching a particular diagnosis. Each diagnosis, and diagnostic situation, has its own fingerprint illustrating factors that are the most challenging and frequent contributors to delayed or misdiagnosis. Prospectively, designing safeguards to protect against these pitfalls and challenges will be needed to dramatically reduce diagnosis errors, but an important first step is awareness of the profile of where and how diagnosis fails for these selected diagnoses.

'Pull' systems in lean quality improvement represent engineering designs that emphasise paving smooth and reliable paths that can help standardise processes and lead to less friction, waste and error. Can we 'pull the trigger' to fire up our diagnosis improvement imaginations, as well as improve our daily clinical cognition by more systematically triggering key diagnoses as Osler advised a century ago? Can 'triggers pull' an enriched sample of cases to facilitate finding, serving up and learning from diagnostic error

supporting clinicians in triggering relevant diagnostic considerations is an avenue that requires more thorough exploration.

However, providing clinicians (and patients) with an even longer differential diagnosis list risks simply overwhelming and/or annoying busy practitioners. As with computerised reminders of other kinds, over-alerting poses the very real risk that clinicians will end up ignoring the suggestions, particularly if they often represent 'false alarms', needlessly slow workflow, or raise impractical suggestions or liability risks.¹⁶ Even if we avoid the cognitive problems of

excessively long diagnostic lists, clinicians will require thoughtful and parsimonious ways of sorting through the probabilities of these different considerations to avoid harm to patients from false positives and overtesting, not to mention the time and anxiety involved in chasing the false leads.

There are no easy answers here. While reassuring patients and allaying needless anxiety remains an important function and responsibility of all practicing clinicians, studies of high reliability organisations demonstrate that a constant state of awareness and worry about 'what could go

cases? Given requisite culture and focus, and natural hunger for clinicians to learn and improve, we think the answer to these questions is 'Yes'.

Acknowledgements Dr Schiff acknowledges the support of the Risk Management Foundation of the Harvard Medical Institutions in his work on diagnostic errors in medicine.

► An additional appendix is published online only. To view this file please visit the journal online (<http://qualitysafety.bmj.com/content/21/2.toc>).

Competing interests None.

Provenance and peer review Commissioned; internally peer reviewed.

Published Online First 30 November 2011

BMJ Qual Saf 2012;21:89–92.
doi:10.1136/bmjqs-2011-000590

REFERENCES

- Singh H, Giardina GD, Forjuoh SN, *et al*. Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ Qual Safety* 2012;21:93–100.
- Singh H, Thomas EJ, Khan MM, *et al*. Identifying diagnostic errors in primary care using an electronic screening algorithm. *Arch Intern Med* 2007;167:302–8.
- Schiff GD, Kim S, Abrams R, *et al*. Diagnosing diagnosis errors: lessons from a multi-institutional. In: Henriksen K, Battles JB, Marks ES, *et al*, eds. *Advances in Patient Safety: From Research to Implementation*. Vol. 2. Rockville MD: U.S. Agency for Research and Quality (AHRQ), 2005.
- Wachter RM. Why diagnostic errors don't get any respect—and what can be done about them. *Health Aff (Millwood)* 2010;29:1605.
- Singh H, Sethi S, Raber M, *et al*. Errors in cancer diagnosis: current understanding and future directions. *J Clin Oncol* 2007;25:5009–18.
- Singh H, Daci K, Petersen LA, *et al*. Missed opportunities to initiate endoscopic evaluation for colorectal cancer diagnosis. *Am J Gastroenterol* 2009;104:2543–54.
- Mull HJ, Shimada MPPS, Nebeker J, *et al*. A review of the trigger literature: adverse events targeted and gaps in detection. 2009. doi:10.4103/0022-3859.57404.
- Mull HJ, Nebeker JR, Shimada SL, *et al*. Consensus building for development of outpatient adverse drug event triggers. *J Patient Saf* 2011;7:66.
- Kaafarani H, Rosen AK, Nebeker JR, *et al*. Development of trigger tools for surveillance of adverse events in ambulatory surgery. *Qual Saf Health Care* 2010;19:425.
- Classen DC, Lloyd RC, Provost L, *et al*. Development and evaluation of the institute for healthcare improvement global trigger tool. *J Patient Saf* 2008;4:169–77.
- Landrigan CP, Parry GJ, Bones CB, *et al*. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010;363:2124–34.
- Franklyn-Miller A, Falvey E, McCrory P. Patient-based not problem-based learning: an Oslerian approach to clinical skills, looking back to move forward. *J Postgrad Med* 2009;55:198.
- 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.
- Schiff GD, Leape LL. *How Can We Make Diagnosis Safer*. Academic Medicine, 2011 In press.
- Schiff G. *Getting Results: From Theory to Practice and Back Again. Getting Results: Reliably Communicating and Acting on Critical Test Results*. Oakbrook Terrace, IL: Joint Commission Resources, 2006:133–49.
- Phansalkar S, Edworthy J, Hellier E, *et al*. A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. *J Am Med Inform Assoc* 2010;17:493–501.
- Weick KE, Sutcliffe KM, Obstfeld D. Organizing for high reliability: processes of collective mindfulness. *Crisis Manage* 2008;3:81–123.
- Reason J. Human error: models and management. *BMJ* 2000;320:768.