Calibrating how doctors think and seek information to minimise errors in diagnosis

Ashley N D Meyer,[1,2] Hardeep Singh[1,2]

[1]Houston Veterans Affairs Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey VA Medical Center, Houston, Texas, USA
[2]Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, Texas, USA

Correspondence to
Dr Ashley N D Meyer, 2002 Holcombe Boulevard (152), Houston, TX 77030, USA; ameyer@bcm.edu

Linked

► http://dx.doi.org/10.1136/bmjqs-2016-005679

CrossMark

Information gathering is a foundational step of the diagnostic process.[1] It is not possible to synthesise clinical information to make a correct diagnosis without adequate data collection related to a patient's history, physical examination, test results or consultations with other clinicians. However, evidence over the last several decades suggests that failures in information gathering are common and feature prominently in analyses of diagnostic errors.[2–7] Many information-gathering failures are related to history taking, including asking the right questions, which is sometimes based on certain cues from the patient.

In this issue of *BMJ Quality and Safety*, Sheringham *et al*[8] used simulated patient vignettes to understand the role that patient characteristics (including demographics and symptomatology) play in physicians' decisions to investigate for possible diagnosis of lung cancer. They found that despite suggestive initial symptoms, general practitioners (GPs) failed to elicit additional key symptoms from 'patients' that would have suggested the need to investigate further through ordering of tests (chest X-rays or CT scans) or referral to a respiratory consultant. GPs were more likely to initiate investigations when they elicited these additional symptoms that the 'patient' had, but did not initially volunteer. The omission of symptom elicitation persisted even in patients with higher risk of lung cancer. While the authors found patient characteristics, such as race and age, related to GPs' investigational behaviours, these factors only accounted for a small proportion of failures to investigate.

What then can explain these differences in information gathering by GPs? In real-world practice, one of the usual suspects is insufficient time,[9] which we ourselves have implicated before.[3]

However, in this study, there was no specific time restriction on the GPs, suggesting we must seek additional reasons. In the real world, physicians sometimes have difficulty in extracting relevant historical data from the abundance of information in electronic health records[10] and thus may not ask the right questions. This again is unlikely to be true for these vignettes. Physicians could also be subjected to cognitive biases[11] and clinical reasoning failures as they juggle uncertainty and balance underdiagnosis on one hand and overzealous diagnostic pursuits on the other. Within the confines of decision-making uncertainty, one concept we described previously[12] which we believe could have an important role to play is lack of calibration, that is, when physicians' confidence in the accuracy of their decisions is not properly aligned with their actual accuracy.[12] We posit that information-gathering failures often relate to *miscalibration of physicians' thinking processes* and that this miscalibration could be a major reason for diagnostic errors and adverse patient outcomes. In this editorial, we highlight how miscalibration can adversely affect information gathering and suggest some next steps.

UNDERSTANDING THE ROLE OF MISCALIBRATION IN DIAGNOSIS

Metacognition refers to thinking about one's own thinking.[13] Our metacognition is miscalibrated when our confidence in our performance of cognitive tasks does not correctly mirror our actual performance. Miscalibration can be hazardous for physicians because self-assessments of their thinking determine additional steps they need to take to solve problems or make decisions, including seeking additional information or help. When it affects diagnostic decision making and

clinical problem solving, it could lead to erroneous diagnoses.

Miscalibration can manifest as either *overconfidence*, where confidence is higher than it should be given performance, or as *underconfidence*, where confidence is lower than it should be given performance. In the realm of diagnostic error, overconfidence has been touted as the more insidious problem.[14] However, both forms of miscalibration can adversely affect information gathering and thus diagnostic decision making. For example, when underconfident, physicians might collect too much data and over investigate through unnecessary tests or referrals, leading to increased healthcare costs.

When physicians are miscalibrated about their decision-making processes in the overconfident direction, they lack awareness about the need to (1) continue seeking information or (2) obtain the help they need to ultimately make correct diagnoses for patients.[12] Both of these situations are associated with diagnostic errors. Thus, when overconfident physicians arrive at (incorrect) diagnoses, they might 'stop short' or fall prey to premature closure in the search for information or for explanations to patients' health problems. Specifically, when overconfident, physicians might curtail questioning about patients' histories or symptoms, they might stop seeking additional information from patients' health records and they might order fewer tests in a failure to properly disconfirm or examine competing hypotheses. In a previous study, we found that even as additional diagnostic information became available (eg, new test results), physicians tended to stay overconfident and not change their diagnoses, suggesting that they may not use the additional information to improve their calibration or diagnostic accuracy.[12]

Miscalibration as overconfidence may also be a precursor to failure to seek additional diagnostic help from text resources, search engines, computerised diagnostic decision support and reaching out to other physicians through either informal consultations or referrals.[12] Despite the availability of these resources, the potential for these resources is often not achieved, partly because of workflow issues[15] or perceptions of looking incompetent in front of patients.[16]

HOW DO PHYSICIANS BECOME BETTER CALIBRATED?

If physicians' miscalibration can adversely affect information acquisition and thus accuracy of diagnoses, what should be next in our scientific pursuit of making physicians better calibrated? While research on this question is very much in its infancy, we propose several avenues for further exploration.

Learning from errors or obtaining feedback has been suggested for improving calibration related to diagnostic decision making[17] and has long been successfully used in the field of cognitive psychology.[18]

Furthermore, additional analyses of our own work on calibration show physicians working in academic environments have better calibration than their non-academic peers.[12] This might be because academicians are involved in more educational activities, continuous learning and feedback through the teaching of medical students, interns, residents and fellows. These activities might force the academic physicians to examine and improve on their own diagnostic decision-making processes, thus fostering better understanding of their performance and how they can improve. In addition, physicians in academic settings might have more exposure to patients with conditions that are 'diagnostic challenges' and might receive more rapid feedback from specialists or results from investigations (especially in inpatient settings) when compared with physicians in community-based or outpatient settings. The latter might receive less structured or delayed feedback of patients' diagnostic outcomes, especially when 'diagnostic challenges' are referred to academic tertiary care centres. Exact mechanisms for the difference between academic and non-academic physicians are conjectural, and additional research is needed on how to develop meaningful feedback and test its effectiveness on calibration.

For example, figuring out how to provide effective, systematic feedback related to diagnostic decision making is a challenge in itself. Research shows that feedback that is timely, given by a trusted source and is actionable is most effective.[19] Determining the content of the feedback to make it actionable might be the biggest hurdle for improving calibration in diagnostic decision making. To make feedback and learning actionable, it might have better utility if aimed at not just physicians' diagnostic *outcome* performance (error, delayed or wrong diagnosis), but rather at diagnostic *process* performance, such as information-gathering performance. It is also unknown if feedback on a specific case will transfer to learning more generally[20] or if feedback will lead to physicians' hypervigilance and as a result, excessive investigations. Nonetheless, the use of simulated patient cases, such as those used in Sheringham *et al*'s study, might make a worthy medium to deliver such feedback when the specifics regarding effective content are determined.

An additional avenue that could be explored for improving physician calibration in diagnosis is the use of diagnostic decision support early on in the data-gathering process. For instance, physicians should be encouraged to consult online diagnostic decision support engines that generate lists of possible diagnoses for consideration based on their patients' symptoms and, ideally, this list could be automatically generated through decision support embedded into the electronic health record. Studies show that diagnostic accuracy improves when physicians use decision support in the early phases of diagnosis where hypotheses are still being formed[21] and this decision

support might improve calibration related to early data gathering and hypothesis formation if used during decision making, rather than as a last verification. In addition, cognitive strategies, such as the 'Red Team/Blue Team Challenge' promoted by the Australian Clinical Excellence Commission,[22] could be tested as methods to improve calibration. This challenge is fashioned after a military exercise in which the 'Red Team' listens and critically challenges the decision-making processes of the 'Blue Team', thereby fostering prevention and awareness of poor decision-making processes.

Physicians must continue to improve their metacognition and clinical reasoning abilities. In addition to these potential solutions above, a culture where errors are openly discussed and a learning environment that supports and trains against the fallibility of human cognition could ensure better calibration.[23] Despite these suggestions, however, increasing knowledge[24] and gaining cumulative experience and expertise through both real and simulated cases might still be the best recipes for improving information gathering and diagnosis.

CONCLUSION

Only with sufficient information gathering can a physician hope to properly interpret and integrate all of a patient's clinical information, determine a working diagnosis, and establish the correct diagnosis and treatment for a patient. Although the oft-hoped-for 1-hour long patient visit might ameliorate some of the difficulties in gathering enough information to make an accurate diagnosis, as the study by Sheringham *et al* suggests, additional time alone is unlikely to be sufficient. Calibration to know when to seek more information or when to seek more help in solving our diagnostic challenges is key for avoiding diagnostic errors and adverse patient outcomes.

REFERENCES

1 National Academies of Sciences Engineering and Medicine. *Improving diagnosis in health care*. Washington DC: The National Academies Press, 2015.
2 Singh H, Giardina TD, Petersen LA, *et al*. Exploring situational awareness in diagnostic errors in primary care. *BMJ Qual Saf* 2012;21:30–8.
3 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.
4 Zwaan L, Thijs A, Wagner C, *et al*. Does inappropriate selectivity in information use relate to diagnostic errors and patient harm? The diagnosis of patients with dyspnea. *Soc Sci Med* 2013;91:32–8.
5 Zwaan L, Thijs A, Wagner C, *et al*. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Acad Med* 2012;87:149–56.
6 Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. *Med Decis Making* 1991;11:233–9.
7 Medford-Davis L, Park E, Shlamovitz G, *et al*. Diagnostic errors related to acute abdominal pain in the emergency department. *Emerg Med J* 2016;33:253–9.
8 Sheringham J, Sequiera R, Myles J, *et al*. Variations in GPs' decisions to investigate suspected lung cancer: a factorial experiment using multimedia vignettes. *BMJ Qual Saf* 2017;26:449–59.
9 Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med* 1999;74 (10 Suppl):S138–43.
10 Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? *N Engl J Med* 2010;362:1066–9.
11 Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775–80.
12 Meyer AN, Payne VL, Meeks DW, *et al*. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med* 2013;173:1952–8.
13 Dunlosky J, Metcalfe J. *Metacognition*. Thousand Oaks, CA, USA: Sage Publications, Inc, 2009.
14 Graber M, Berner EE. Diagnostic error: is overconfidence the problem? *Am J Med* 2008;121:S1–S46.
15 Miller A, Moon B, Anders S, *et al*. Integrating computerized clinical decision support systems into clinical work: A meta-synthesis of qualitative research. *Int J Med Inform* 2015;84:1009–18.
16 Shaffer VA, Probst CA, Merkle EC, *et al*. Why do patients derogate physicians who use a computer-based diagnostic support system? *Med Decis Making* 2013;33:108–18.
17 Croskerry P. The Feedback Sanction. *Acad Emerg Med* 2000;7:1232–8.
18 Lichtenstein S, Fischhoff B. Training for calibration. *Organ Behav Hum Perform* 1980;26:149–71.
19 Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Med Care* 2009;47:356–63.
20 Goodman JS, Wood RE, Chen Z. Feedback specificity, information processing, and transfer of training. *Organ Behav Hum Decis Process* 2011;115:253–67.
21 Kostopoulou O, Rosen A, Round T, *et al*. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. *Br J Gen Pract* 2015;65:e49–54.
22 Take 2-Think, Do. *Clinical Excellence Commission*. 2015. https://c.ymcdn.com/sites/improvediagnosis.site-ym.com/resource/resmgr/Take_2_-_BThink_Do_clinician.pdf (accessed 28 Aug 2016).
23 Singh H, Schiff GD, Graber ML, *et al*. The global burden of diagnostic errors in primary care. *BMJ Qual Saf* 2017;26:484–94.
24 Dhaliwal G. Premature closure? Not so fast. *BMJ Qual Saf* 2017;26:87–9.