ORIGINAL RESEARCH

OPEN ACCESS

Improving diagnostic performance through feedback: the Diagnosis Learning Cycle

Carolina Fernandez Branson,[1] Michelle Williams,[2] Teresa M Chan [ID],[3] Mark L Graber,[4] Kathleen P Lane [ID],[5] Skip Grieser,[6] Zach Landis-Lewis,[2] James Cooke,[2,7] Divvy K Upadhyay,[8] Shawn Mondoux [ID],[3] Hardeep Singh [ID],[9] Laura Zwaan [ID],[10] Charles Friedman,[2] Andrew P J Olson [ID] [5,11]

Check for updates

ABSTRACT

Background Errors in reasoning are a common cause of diagnostic error. However, it is difficult to improve performance partly because providers receive little feedback on diagnostic performance. Examining means of providing consistent feedback and enabling continuous improvement may provide novel insights for diagnostic performance.

Methods We developed a model for improving diagnostic performance through feedback using a six-step qualitative research process, including a review of existing models from within and outside of medicine, a survey, semistructured interviews with individuals working in and outside of medicine, the development of the new model, an interdisciplinary consensus meeting, and a refinement of the model.

Results We applied theory and knowledge from other fields to help us conceptualise learning and comparison and translate that knowledge into an applied diagnostic context. This helped us develop a model, the Diagnosis Learning Cycle, which illustrates the need for clinicians to be given feedback about both their confidence and reasoning in a diagnosis and to be able to seamlessly compare diagnostic hypotheses and outcomes. This information would be stored in a repository to allow accessibility. Such a process would standardise diagnostic feedback and help providers learn from their practice and improve diagnostic performance. This model adds to existing models in diagnosis by including a detailed picture of diagnostic reasoning and the elements required to improve outcomes and calibration.

Conclusion A consistent, standard programme of feedback that includes representations of clinicians' confidence and reasoning is a common element in non-medical fields that could be applied to medicine. Adapting this approach to diagnosis in healthcare is a promising next step. This information must be stored reliably and accessed consistently. The next steps include testing the Diagnosis Learning Cycle in clinical settings.

INTRODUCTION

Diagnostic error is a significant cause of morbidity and mortality,[1] and flaws in clinical reasoning contribute to a substantial proportion of these errors.[2] There are many factors that lead to suboptimal clinical reasoning,[3] and interventions have been proposed to reduce them.[4] However, few interventions have been shown to improve diagnosis in practice,[5] and what constitutes or delineates the boundaries of clinical reasoning is highly variable and not standardised.[6] Thus, designing interventions to improve clinical reasoning with the aim of improving diagnostic performance remains a challenging task.

One key factor inhibiting the improvement of clinical reasoning is physicians' lack of knowledge about diagnostic outcomes that could inform and improve future clinical decisions. The culture of medicine generally is one of 'no news is good news'; physicians in practice may rarely find out if their diagnostic reasoning was incorrect. Thus, clinicians may remain overconfident and poorly calibrated in their diagnostic reasoning and/or clinical decision-making.[7][8] Furthermore, diagnosis is more than simply a label for a patient's health condition—it is also the process by which clinicians gather and synthesise clinical information.[9] This process reflects a complex mix of analytic and non-analytic cognitive processes on which feedback is important for improving performance. Further, clinical reasoning is context-dependent, and thus providing specific feedback about specific cases is much more likely to be effective than general improvement strategies.

Calibration is defined as the relationship between the accuracy of a decision and an individual's perception of, and confidence in, that accuracy.[10] In a clinical context, being well calibrated means that providers neither underestimate the

accuracy and/or quality of a decision or process used to make a decision (which may lead to overtesting and/or overuse) or overestimate the accuracy and/or quality of a decision or process used to make a decision (which may lead to inaccurate conclusions or premature closure[11]).

Consistent feedback is one method proposed to improve diagnostic performance and improve calibration.[7 10] Feedback is a key component to deliberate practice, which has been shown to be an effective method to achieve mastery-level performance in a variety of fields.[12] However, clinicians receive relatively little formal feedback on their diagnostic decision-making.[1] Part of the reason for this lack of feedback is that information about the diagnostic process is challenging to capture, patients' subsequent clinical outcomes are not well tracked over time, and transitions of care between providers are ubiquitous in modern healthcare. Thus, there is no systematic feedback loop that compares initial diagnostic hypotheses, diagnostic processes and later diagnoses. Without outcomes-informed reflection on the process of diagnosis, it remains difficult to identify how clinical reasoning, and ultimately diagnosis, can be improved in concrete and tangible ways. This is further compounded by the ubiquity and complexity of uncertainty in medicine, making follow-up and reflection fundamental. However, other fields of human performance have well-developed feedback loops and well-calibrated experts due to this continuous feedback. Engaging with experts from these fields to learn about their methods for continuous feedback and improvement may lend promising insights.

In this paper, we describe the development of a new model developed through engagement with experts and literature both within and outside medicine. The new model aims to fill gaps in the literature on improving diagnostic performance through outcomes-informed feedback about decision-making and is ripe for testing in the clinical environment.

METHODS

The development of the model was the result of a six-phase qualitative research process. First, we conducted a literature review to identify relevant existing models of diagnostic performance and how feedback contributes to learning and improvement. This included a search of PubMed as well as identification of possible relevant models referenced in these articles, some of which were specific to medicine and some were not.

Second, we performed a hypothesis-generating, exploratory survey (online supplemental appendix 1) designed to identify which fields outside of medicine shared similar task elements or cultural characteristics with medicine so that we could interview individuals in these professions and learn the methods of feedback they used for performance improvement in the course of their work. The survey was developed by the study team and then refined after cognitive interviews

were performed with practising clinicians and medical educators in a works-in-progress session. We asked targeted questions about the types of tasks these professions perform, the 'stakes' involved, and the return on investment (ROI) for performance improvement, among others, they considered most similar to medicine. In general, we assumed that improving the practice of diagnosis in medicine primarily benefits the patient (rather than the clinician), clinician diagnostic performance improvement is intrinsically motivated rather than externally incentivised, the 'stakes' are high, and that the effects of improvements on performance have somewhat unpredictable improvements in health outcomes since interventions designed to improve clinician performance often do not link to outcomes. We allowed respondents to choose from our list or fill in professions they thought would fit but were not on the list. We administered the survey to health sciences professionals, including clinicians focused on diagnosis identified through the Society to Improve Diagnosis in Medicine, as well those who have expertise in diagnostic safety and performance review. We provided them with a list of a wide variety of professions to consider for inclusion in the interviews. Although these health science professionals may not be intimately familiar with the tasks of other fields, their opinions gave us a starting point to identify which professions to target for indepth interviews. Responses were tabulated and professions selected for the next stage of the project based on representation across multiple domains of the survey by multiple respondents.

Next, based on these results, we conducted semistructured interviews (online supplemental appendix 2) of professionals from the fields outside medicine as well as practising clinicians and medical educators. Individuals were identified by the primary study team as well as from referrals from other interviewees. The interview script was developed by the primary study team. The script for these semistructured interviews remained the same throughout the interviews. The interviews were conducted by a member of the research team and recorded via Zoom and transcribed. Interviews were analysed thematically for content using NVivo (Vermont, USA). The first five interviews were analysed by two members of the research team and the codebook iteratively refined; once excellent agreement was achieved, the remainder of the interviews were analysed by one team member.

Based on the existing models, surveys and interviews, we developed a draft of our model, called the Diagnosis Learning Cycle (figure 1). The next phase of the project was to convene a meeting of experts in medicine and the three other targeted fields we identified to share best practices in feedback and performance improvement and to get feedback on our preliminary model (see online supplemental appendix 3 for a list of attendees). In January 2020, we convened a 2-day meeting in St

A: Diagnosis Learning Cycle
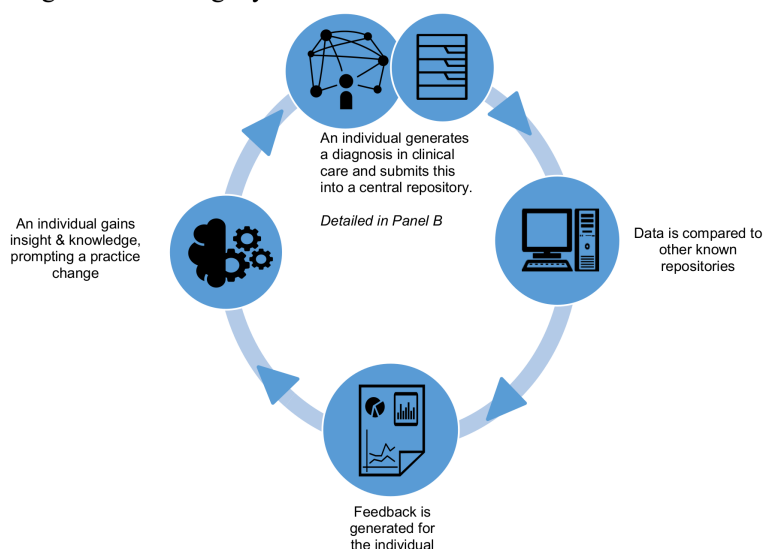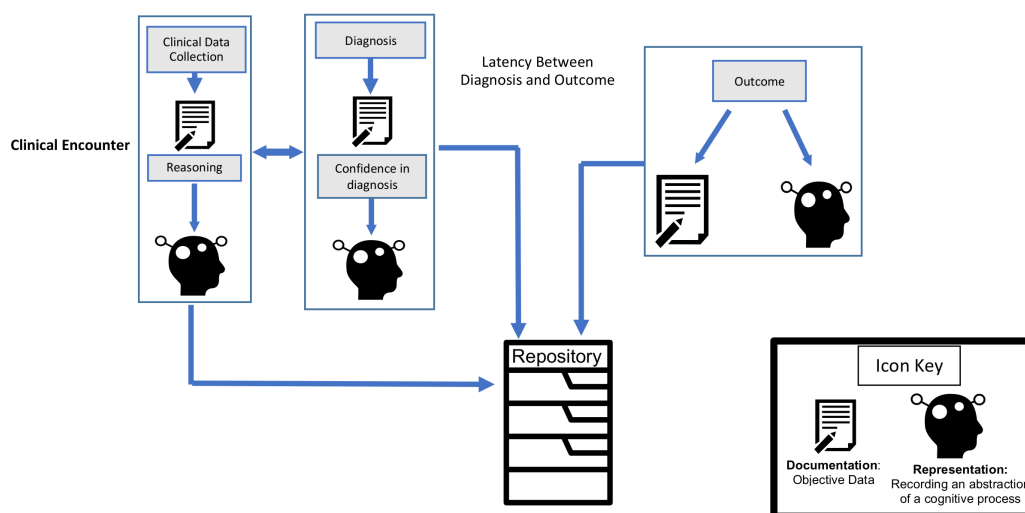

B: Diagnosis Process Capture

Figure 1 The Diagnosis Learning Cycle (A) and the Diagnostic Process Capture (B). (A) The Diagnosis Learning Cycle includes the Diagnosis Process Capture from which outcomes are placed into a repository. After a threshold (number of cases, period of clinical service) is met, the outcomes in the repository are accessed to allow comparison, by which a provider compares their initial diagnostic hypotheses with associated outcomes. After comparison, providers receive feedback by interacting with an external resource (such as a coach, colleague or information resource), thereby modifying and improving future knowledge and performance. This cycle is iterative over time. (B) The Diagnosis Process Capture represents how information in a single case is captured and placed into the repository to allow the Diagnosis Learning Cycle to proceed. Information is captured, in general, in two ways: clinical data (such as vital signs, laboratory tests and physical examination findings) are *documented* (pen and paper icon), while cognitive processes (eg, differential diagnosis) and attitudes (eg, confidence) are *represented and recorded* (silhouette icon). These initial factors are then compared with the final (or intermediate) clinical outcomes and associated cognitive processes. The initial decisions and information are placed in the repository along with the intermediate/final outcomes, enabling better calibration and improved performance.

Petersburg, Florida. Feedback was collected during the convening meeting, and subsequently the primary research team refined the model into its final form.

RESULTS
Existing models
We identified two models within medicine that outline the diagnostic process and calibration and two models

outside of medicine that theorise how individuals learn and improve task performance. The models from within medicine provided a starting point for thinking about the diagnostic process, feedback loops and calibration. The *Safer Dx framework for the measurement and reduction of diagnostic errors*[13] is a model that includes process dimensions, measurement of diagnostic errors and links to improved outcomes. This

model captures the general sequence of, and relationship between, events in the diagnostic process, but does not provide explicit details for analysis of clinicians' decision-making. Croskerry's[14] model emphasises the fundamental importance of knowing the outcomes of decisions made in enabling calibration among physicians. Similar to the Safer Dx model, it does not provide granularity necessary to operationalise this feedback.

Two models from outside of medicine helped us to better characterise learning and performance improvement. Kluger and DeNisi's[15] work on the feedback intervention (FI) theory on the effects of FIs on task-learning processes is valuable because it illustrates the process of trial and error, as well as how knowing outcomes contributes to different learning effects, and ultimately improving performance. However, this model is not about diagnostic reasoning specifically and thus does not capture the nuances nor the context of the diagnostic process.

Finally, Ericsson and Harwell's model[16] shows how musicians can improve performance by listening to representations of ideal ways to execute performance as well as listen to their current performance in order to compare the two and achieve desired performance. This model influenced our use of the term 'representation' (external recordings of cognitive processes) and sparked our interest in the concept of 'comparison' (when initial decisions and outcomes are examined in light of one another). It also helped us posit what types of representations that could be captured in the diagnostic process so that they could be compared and used as a tangible source of feedback to improve physician performance.

Taken together, these models provide important insights into the diagnostic process and also how individuals learn and improve performance based on feedback. However, there remained gaps with respect to specific strategies for learning and improvement in diagnosis.

Exploratory survey
Twenty-three individuals were invited to complete the survey and 13 responded (57%). Eight professions were identified as possessing similar tasks and/or other characteristics. These included meteorology, agriculture, air traffic control, aviation, clergy, team sports, law enforcement and music. For example, in meteorology and sports, improving performance may have a large influence on outcomes, similar to medicine. Air traffic controller, pilot and law enforcement were rated as similar to medicine in that the 'stakes', human lives, were high. Finally, law enforcement, the clergy and music are all professions where the ROI is often unclear, as it is in medicine.

Semistructured interviews
We conducted semistructured interviews with individuals from these eight professions as well as several practising clinicians with an interest in quality and safety so that we could better compare tasks, processes and cultures of these other professions with medicine. In all, we conducted 18 semistructured interviews with 9 individuals who were practising in medicine and 9 individuals who represented other professions. Through analysis of the interviews, we determined meteorology, aviation and team sports each had the most task-similar scenarios as physicians, but different enough from each other that each one would provide a unique learning opportunity. We thus engaged more with professionals in these fields throughout the project.

Similar to medical professionals, meteorologists have access to large amounts of data to help facilitate decision-making and human lives are at stake. In addition, meteorology has been proposed as sharing similarities with physicians because both possess task uncertainty (ie, forecasting and diagnosis). However, unlike medicine, meteorologists have a much more clear picture of weather outcomes, as outcomes of predicted weather events are mostly known. The aviation field is similar to medicine in that there may be high-stakes consequences of making the wrong decision. Further, aviation has been a leader in driving safety initiatives, and the movement to improve safety in healthcare has been stimulated by, and modelled on, advances in aviation safety.[17]

Because aviation is now one of the safest industries, we hoped to learn strategies that could translate to improving safety in diagnosis. Finally, in team sports, feedback is ubiquitous and we could learn from the advanced types of strategies coaches use to prepare for a game and apply this knowledge to develop strategies for improvement in diagnosis.

Preliminary development of the model, consensus conference and further refinement
Based on these findings, the primary research team developed a model to describe how diagnostic decision-making can improve over time based on feedback about clinical practice. The meeting hosted 22 professionals from a variety of fields, including medicine/medical education (physicians, medical educators and diagnosis researchers), aviation, meteorology and athletics from the USA, Canada and Europe. Six of these individuals were familiar with the project as they had participated in the survey or a semistructured interview, 5 were members of the core study team and 11 were new to the project. The convening meeting consisted largely of structured discussions designed to refine the model. The primary research team then continued to refine and develop the model so that it was easy to understand visually and contained the appropriate components. In the next section, we will explain the new model that was generated from our research: the Diagnosis Learning Cycle.

Table 1 Comparison of medicine with other fields of performance with respect to decisions, outcomes, documentation of data and representations of mental processes as well as how feedback occurs

Field	Decision	Outcome	Documented data	Representations of mental processes	How feedback about decision-making processes and outcomes occurs
Air traffic control	Can a specific aircraft safely take off or land at a given time in the context of weather and traffic conditions?	Maintenance of safe distance between aircrafts at all times. Safe take-offs and landings.	Radar data recordings that depict aircraft flight paths, altitudes, speeds, and runway activity.	Recordings of all verbal communications between pilots and air traffic controllers, including controller instructions.	Checklists and debriefings of simulator training, on-the-job training, and error investigations.
Meteorology	Should a specific weather warning or alert be issued?	Advance warning about severe weather without excessive false positive warnings.	Model predictions, radar maps at given points in time, actual recordings of severe weather events, storm spotter reports, storm surveys.	Chat transcripts, recordings of conversations and consultations with other experts.	Consistent review of verification data, tracking of false positive and false negative rates.
Team sports (eg, football)	Should a specific play be used at a given time?	Success of an individual play at a given point, overall win/loss.	Recordings of the game/practice.	Notes from coaching staff.	Consistent review of film with debriefing after practice and games.
Medicine	What is the correct prioritised differential diagnosis for this patient?	Accurate, timely and efficient diagnosis (often unknown).	Notes in the medical record, imaging, laboratory data.	Sometimes contained in clinician notes.	Occurs uncommonly through formal peer review, informal discussion or personal methods to follow up on cases.

The Diagnosis Learning Cycle

We learnt that the fields outside of medicine that we engaged each had detailed external representations, or 'information artefacts' of their hypotheses, decision-making processes and outcomes. These artefacts make comparison between hypotheses, processes and outcomes possible, thus enabling learning. In discussing these artefacts, we determined there were two main forms: *documentation* (ie, recording data, such as temperature, speed, distance or symptoms) as well as *representations of mental processes* (ie, information artefacts about decision-making steps, such as reasoning and hypothesis generation). We learnt that the other professional fields not only documented data but had some form of externally represented information artefacts that captured these more complex processes of decision-making. This was often done through transcripts or written notes. The characteristics of each of the included fields with respect to each of these factors are shown in table 1. Medicine appears to have a gap in recording complete representations of mental processes, diagnostic outcomes and the consistent delivery of feedback.

Ultimately, we learnt each profession had information artefacts that were external representations of mental processes available. In medicine, diagnostic hypotheses, processes and outcomes are frequently poorly represented, often not compared, and confidence in a diagnosis is rarely captured. In many ways, initial diagnostic hypotheses, diagnostic processes and diagnostic outcomes are separated so substantially that their comparison is the exception rather than the norm.

The Diagnosis Learning Cycle (figure 1A) represents an idealised, yet pragmatic, means by which individuals and/or teams can improve their diagnostic performance over time. The cycle begins with the diagnosis process capture (figure 1B), which involves recording both initial diagnosis and outcomes data, as well as proposes developing information artefacts that feature representations of decision-making processes and confidence in diagnostic decisions. This information is placed into a repository (in the Electronic Health Record (EHR) or another electronic source) which is then accessed when a certain condition is met (such as a number of cases, a period of service or a predetermined major discrepancy between diagnosis and outcome) and compared. Comparison involves examining the initial data as well as the diagnostic hypothesis, diagnostic process, confidence and outcomes. It must be noted that outcomes can exist both as documentation of data (result of a diagnostic test) or as a representation of a mental process (ie, a clinician's final diagnosis). After comparison, an individual or team interacts with an information resource (eg, another person, educational resource, etc) to enable meaning-making of the comparison. This then ideally leads to knowledge modification, and ultimately incorporation and improvement of future practice.

As is shown in the Diagnostic Process Capture, certain parts of the diagnostic process are already

documented: including collecting data on patient symptoms and documenting a diagnosis in the EHR. What is currently missing is the information artefact that is an external *representation* of professionals' clinical reasoning processes and their confidence in the decisions made. These two are much more abstract—and likely variable among providers—than documenting symptoms or diagnoses, but just because they are abstract does not mean they cannot be represented externally. Our model proposes representing clinical reasoning and confidence externally in the form of an information artefact so they may be documented in a system, similar to the way symptoms and diagnoses now are. By doing this, clinicians will have a more clear, comprehensive picture of the information and reasoning that went into a diagnosis, thus allowing them to learn from the process by using accurate information, rather than relying on recall alone.

In order to be meaningful, the representations in the Diagnostic Process Capture must have key attributes, including fidelity (how well they represent the context, decision and outcome); persistence (how durable the representations are over time); and specificity (to what degree the outcome is determined by the decision). This Diagnostic Process Capture cycle continues to iterate until some condition is met (number of cases, end of an activity, etc) and then feeds into the Diagnosis Learning Cycle, as illustrated above. Thus, an individual or team could choose to access the repository to compare diagnostic hypotheses and processes, confidence and outcomes after a period of time or accumulation of cases; this likely will be a period of clinical service, a given number of shifts or a period of time (eg, monthly).

DISCUSSION

We put forth a new model that describes how clinicians can learn and improve about diagnostic reasoning in the modern healthcare environment through feedback about their decision-making. Our model complements the existing literature because it is a model specific to diagnosis that offers concrete recommendations for what types of information to capture and give as feedback to providers, theorises the diagnostic learning process, and propels concepts about diagnosis and learning into an actionable model. The diagnosis learning model adds to the Safer Dx framework by proposing a detailed process of how to create data that will be used to fill the existing gap between initial diagnostic assessment and tracking outcomes. By capturing confidence and reasoning and proposing these elements be stored in a repository, we are adding the concept of comparison to the Safer Dx model, which we believe will help to enhance diagnostic learning and ultimately safety. Similarly, we add to Croskerry's[14] model by outlining the details of the decision-making process and what should be captured during that process to enable better calibration. Ultimately,

our work adds to the medical literature on diagnostic decision-making by adding a level of detail about what elements of decision-making should be captured to enhance diagnostic learning that are currently missing.

We used models that theorise learning and performance improvement outside of medicine to help us conceptualise how the elements of hypothesis generation, representation and comparison contribute to diagnostic improvement and learning. We apply the elements of both Kluger and DeNisi's[15] model of how feedback influences learning (where we incorporate hypothesis generation and comparison) and Ericsson and Harwell's model[16] (where we consider the importance of generating representations and comparing them) to the diagnostic realm. Thus, we combine elements of these models and place them into a new model that allows for a specific application of these concepts amenable to testing in the clinical setting.

This model outlines a clear process for diagnostic learning, identifies the data that need to be captured and offered to providers as feedback, and applies concepts about learning and performance improvement from other fields to medicine. By introducing the concept that clinical reasoning (in the form of initial hypotheses, confidence and diagnostic outcomes) needs to be externally *represented* and provided to healthcare professionals in the form of an information artefact to enable comparison, we are enabling a standardised process of diagnostic learning and improvement to emerge.

Testing this new model (ie, testing how to best externalise and preserve clinical reasoning and outcomes for comparison) can be achieved in a variety of ways by considering a multiplicity of factors that contribute to the success and challenges of implementation.[18] The experts from our convening concluded that testing the feasibility of this model could begin with certain symptoms and associated diagnoses, such as paediatric abdominal pain (appendicitis, constipation, urinary tract infection), adult chest pain (aortic dissection, pulmonary embolism, acute coronary syndrome) or loss of consciousness (syncope, seizure). Relevant symptoms and diagnoses will vary across specialties and care settings. Another theme that emerged from the convening is that it is critical that collecting and retrieving this information fit within the healthcare workflow so time and pressure would not be added to providers' already time-crunched schedules. Recording 'think-alouds' while the patient is in the room, or after immediately leaving the room, was one suggestion for how reasoning and confidence could be captured in a working diagnosis without adding more work to a clinician's schedule.

Capturing diagnostic outcomes is challenging and likely also context-dependent, and implementation must consider this.[19] For example, some care settings, such as emergency medicine and urgent care, are primarily focused on not missing severe or

life-threatening conditions versus absolute diagnostic accuracy. Further, short-term diagnoses (within a few days of presentation) are different from long-term diagnoses (months to years after presentation)—diagnosis evolves over time, and receiving short-term feedback (eg, during the first few days of a hospitalisation) may provide different insights from long-term feedback (eg, based on diagnoses months to years after an encounter). Feedback based on longer-term outcomes is likely to be more logistically challenging. Clinicians may be less likely to remember specific details if a long time has elapsed between decision and outcome; however, developing reliable means to capture representations will obviate some of this difficulty. In order to compare the hypothesis with the outcome, a software system could be created where this information could be easily accessed by providers so they could see if their diagnostic hypothesis matched the outcome. Testing this model may also look different based on provider specialty and condition; therefore, the model is constructed to be flexible and amenable for use for a variety of specialties.

Limitations

Our model has several limitations, including that it has yet to be tested. We hypothesise that capturing reasoning, initial diagnoses and outcomes and comparing them with one another will improve diagnostic performance, but this has neither been tested nor proven in medicine. Another important limitation is that dissecting the clinical reasoning process as illustrated in the model may be overly simplistic in relation to how diagnosis works in reality, as multiple people may be involved in diagnostic reasoning on a single case, and other systems-related factors are at play. This is further complicated by the context in which diagnosis is made, including patient interactions, mental state of the provider and many others. Similarly, the underlying process that an individual uses to make a diagnosis in a case often involves nonanalytic reasoning that would be challenging to represent accurately.[20] Further, our model mostly addresses individual decision-making and improvement and does not take into account other important theories that are important for clinical reasoning and diagnosis, such as situated cognition or other social cognitive theories.[21] Similarly, we did not formally address the use of artificial or augmented intelligence, which is emerging as an important consideration and adjunct in modern clinical reasoning. Lastly, much of the success of this model relies on accurate and thoughtful clinical documentation; in the USA, much of the major driving force shaping clinical documentation is related to billing and coding, thus leaving actual clinical reasoning obfuscated.

CONCLUSION

In conclusion, the Diagnosis Learning Cycle model was created using a six-phase qualitative research process that included learning from other professions and using existing models. This model adds to the literature by combining features of FI and learning theory with a close analysis of the diagnostic process, and applies the concept of externalising mental representations and comparison to diagnosis. Our model adds to the literature by taking portions of the clinical reasoning process such as reasoning and confidence from abstraction into the realm of representation by externalising them as information artefacts. By adding a repository to clinical practice where hypotheses, confidence and outcomes may be compared, this model enables the provision of regular feedback to clinicians on their diagnostic performance. We envision that this model can enhance diagnostic learning and improve diagnostic performance. The next steps will include testing the feasibility of the model across a number of contexts and further refining it to enable learning and improvement.

Author affiliations
[1]School of Communication Writing and the Arts, Metropolitan State University, St. Paul, Minnesota, USA
[2]Learning Health Sciences, University of Michigan, Ann Arbor, Michigan, USA
[3]Emergency Medicine, McMaster University, Hamilton, Ontario, Canada
[4]Society to Improve Diagnosis in Medicine, Chicago, Illinois, USA
[5]Medicine, University of Minnesota Medical School, Minneapolis, MN, USA
[6]Colorado State University, Fort Collins, CO, USA
[7]Family Medicine, University of Michigan, Ann Arbor, Michigan, USA
[8]Geisinger, Danville, Pennsylvania, USA
[9]Center for Innovations in Quality, Effectiveness and Safety, Michael E DeBakey VA Medical Center and Baylor College of Medicine, Houston, TX, USA
[10]Institute of Medical Education Research Rotterdam, Erasmus MC, Rotterdam, The Netherlands
[11]Pediatrics, University of Minnesota Medical School, Minneapolis, MN, USA

Twitter Teresa M Chan @TChanMD, Zach Landis-Lewis @zachll, Hardeep Singh @HardeepSinghMD, Laura Zwaan @laurazwaan81 and Andrew P J Olson @andrewolsonmd

disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Teresa M Chan http://orcid.org/0000-0001-6104-462X
Kathleen P Lane http://orcid.org/0000-0003-3125-3273
Shawn Mondoux http://orcid.org/0000-0002-6752-3485
Hardeep Singh http://orcid.org/0000-0002-4419-8974
Laura Zwaan http://orcid.org/0000-0003-3940-1699
Andrew P J Olson http://orcid.org/0000-0002-7226-5783

REFERENCES

1 Balogh EP, Miller BT, Ball JR. *Improving diagnosis in health care*, 2016.

2 Graber ML, Rencic J, Rusz D, *et al*. Improving diagnosis by improving education: a policy brief on education in healthcare professions. *Diagnosis* 2018;5:107–18.

3 Scott IA. Errors in clinical Reasoning: causes and remedial strategies. *BMJ* 2009;338:b1860.

4 Ruedinger E, Olson M, Yee J, *et al*. Education for the next frontier in patient safety: a longitudinal resident curriculum on diagnostic error. *Am J Med Qual* 2017;32:625–31.

5 Graber ML, Kissam S, Payne VL, *et al*. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012;21:535–57.

6 Young M, Thomas A, Lubarsky S, *et al*. Drawing boundaries: the difficulty in defining clinical Reasoning. *Acad Med* 2018;93:990-995.

7 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121:S2–23.

8 Meyer AND, Payne VL, Meeks DW, *et al*. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med* 2013;173:1952–8.

9 Hautz WE. When I say… diagnostic error. *Med Educ.* 2018.

10 Meyer AND, Singh H. The path to diagnostic excellence includes feedback to calibrate how clinicians think. *JAMA* 2019;321:737.

11 Zwaan L, Hautz WE. Bridging the gap between uncertainty, confidence and diagnostic accuracy: calibration is key. *BMJ Qual Saf* 2019;28:352–5.

12 Pool R, Ericcson AK. *Peak: secrets from the new science of expertise*. First Mariner Books, 2017.

13 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.

14 Croskerry P. The feedback sanction. *Acad Emerg Med* 2000;7:1232–8.

15 Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996;119:254–84.

16 Ericsson KA, Harwell KW. Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: why the original definition matters and recommendations for future research. *Front Psychol* 2019;10:2396.

17 Helmreich RL. On error management: lessons from aviation. *BMJ* 2000;320:781–5.

18 Greenhalgh T, Wherton J, Papoutsi C, *et al*. Beyond adoption: a new framework for theorizing and evaluating Nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res* 2017;19:e367.

19 Meyer AND, Upadhyay DK, Collins CA, *et al*. A program to provide clinicians with feedback on their diagnostic performance in a learning health system. *Jt Comm J Qual Patient Saf* 2021;47:120–6.

20 Norman GR, Monteiro SD, Sherbino J, *et al*. The causes of errors in clinical Reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med* 2017;92:23-30.

21 Olson APJ, Durning SJ, Fernandez Branson C, *et al*. Teamwork in clinical reasoning - cooperative or parallel play? *Diagnosis* 2020;7:307–12.