International Journal for Quality in Health Care, 2020, 31(8), G97–G102
doi: 10.1093/intqhc/mzz082
Advance Access Publication Date: 28 October 2019
Article

Article

Isolating red flags to enhance diagnosis (I-RED): An experimental vignette study

COREY CHARTAN[1], HARDEEP SINGH[2], PARTHASARATHY KRISHNAMURTHY[3,4], MOUSHUMI SUR[1], ASHLEY MEYER[2], RIAD LUTFI[5], JULIE STARK[5], and SATID THAMMASITBOON[1,6]

[1]Department of Pediatrics, Section of Critical Care Medicine, Texas Children's Hospital/Baylor College of Medicine, 6651 Main St, Houston, TX, 77030, USA, [2]Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey Veterans Affairs Medical Center and the Department of Medicine, Baylor College of Medicine, 2450 Holcombe Blvd Suite 01Y, Houston, TX, 77021 , USA, [3]Department of Marketing and Entrepreneurship, C.T. Bauer College of Business, University of Houston, 334 Melcher Hall, Houston, TX, 77204, USA, [4]Department of Anesthesiology and Pain Medicine, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555,USA and Department of Pediatrics, Baylor College of Medicine, 6651 Main St, Houston TX, 77030, USA, [5]Riley Children's Hospital, Indiana University School of Medicine, 705 Riley Hospital Dr, Indianapolis, IN, 46202, USA, and [6]Center for Research, Innovation and Scholarship in Medical Education and the Department of Pediatrics, Section of Critical Care Medicine, Texas Children's Hospital/Baylor College of Medicine, 6621 Fannin St, Suite A118 Houston, TX, 77030, USA

Address reprint requests to: Satid Thammasitboon, Center for Research, Innovation and Scholarship in Medical Education and the Department of Pediatrics, Section of Critical Care Medicine, Texas Children's Hospital/Baylor College of Medicine, 6651 Main St, Houston, TX 77030, Tel: (832) 826-6230; Fax: (832) 826-4252; USA. E-mail: satidt@bcm.edu

Editorial Decision 24 June 2019; Accepted 24 June 2019

Abstract

Objective: To investigate effects of a cognitive intervention based on isolation of red flags (I-RED) on diagnostic accuracy of 'do-not-miss diagnoses.'

Design: A 2 × 2 randomized case vignette-based experiment with manipulation of I-RED strategy between subjects and case complexity within subjects.

Setting: Two university-based residency programs.

Participants: One-hundred and nine pediatric residents from all levels of training.

Interventions: Participants were randomly assigned to the I-RED vs. control group, and within each group, they were further randomized to the order in which they saw simple and complex cases. The I-RED strategy involved an instruction to look for a constellation of symptoms, signs, clinical data or circumstances that should heighten suspicion for a serious condition.

Main Outcome Measures: Primary outcome was diagnostic accuracy, scored as 1 if any of the three differentials given by participants included the correct diagnosis, and 0 if not. We analyzed effects of I-RED strategy on diagnostic accuracy using logistic regression.

Results: I-RED strategy did not yield statistically higher diagnostic accuracy compared to controls (62 vs. 48%, respectively; odd ratio = 2.07 [95% confidence interval, 0.78–5.5], $P = 0.14$) although participants reported higher decision confidence compared to controls (7.00 vs. 5.77 on a scale of 1 to 10, $P < 0.02$) in simple but not complex cases. I-RED strategy significantly shortened time to decision (460 vs. 657 s, $P < 0.001$) and increased the number of red flags generated (3.04 vs. 2.09, $P < 0.001$).

Conclusions: A cognitive strategy of prompting red flag isolation prior to differential diagnosis did not improve diagnostic accuracy of 'do-not-miss diagnoses.' Given the paucity of evidence-based

Published by Oxford University Press in association with the International Society for Quality in Health Care 2019.

solutions to reduce diagnostic error and the intervention's potential effect on confidence, findings warrant additional exploration.

Key words: diagnostic error, diagnostic accuracy, case complexity, system two, cognitive load, patient safety

Introduction

Diagnostic errors are increasingly described as contributors to preventable morbidity and mortality [1–5]. The 2015 National Academies of Sciences, Engineering and Medicine (NASEM) report, 'Improving Diagnosis in Health Care,' highlights that most people will experience at least one diagnostic error in their lifetime, sometimes with devastating consequences [1]. A population-based estimate suggests that diagnostic errors affect at least 1 in 20 US adults [6]. Novel interventions are needed to address cognitive and system-related factors that lead to these errors [7]. Several reviews have recognized the absence of cognitive interventions and recommended next steps, including rigorously evaluating well-conceptualized and widely endorsed interventions previously suggested in the literature [8, 9].

Clinical reasoning, i.e. 'the cognitive process that is necessary to evaluate and manage a patient's medical problem,' is an essential component of clinical competence [9]. Cognitive errors may result from various factors including inadequate knowledge or faulty data gathering, inaccurate clinical reasoning or faulty verification [8]. Several cognitive contributions to diagnostic error have been studied by multidisciplinary researchers [6], and a large number of strategies for reducing cognitive errors have been suggested. However, most cognitive interventions have not been tested experimentally or in real clinical settings [9–11]. A majority of these cognitive strategies are theory-based approaches and/or pedagogical recommendations [11].

Suggested cognitive strategies to reduce diagnostic errors are largely related to increasing knowledge, improving clinical reasoning and using decision support tools or second opinions. For example, certain recommendations focus on improving clinical reasoning using diagnostic checklists by promoting mindful reflection and a way to improve clinical reasoning [12, 13]. Complex cases that have several possible diagnoses and/or involve potentially serious diagnoses pose special diagnostic challenges. Key symptoms may get overlooked, predisposing these complex conditions to increased risk of diagnostic errors [14, 15]. In fact, recent studies show that diagnostic errors occur despite the presence of clear 'red flags' that warrant further patient evaluation and that differential diagnosis is absent from many error cases [16].

We hypothesized that a prompt for clinicians to isolate red flags among the many clinical findings can facilitate more effective hypothesis generation and lead to more accurate diagnoses. Although the term 'red flag' is used commonly to indicate any information that alarms practitioners to investigate further, the term may mean differently to different people. We created the operational definition of a red flag a priori to provide clarity for this study design and intervention and to explain potential implications of the results. We defined a red flag as a constellation of symptoms, signs, clinical data or circumstances, conceptualized by an individual clinician, that should lead to heightened suspicion for a serious condition and trigger additional evaluation (Fig. 1). It refers to an idea generated de novo with personal meaning to each clinician and for each patient and could be broader than traditional red flags defined as rules of thumb (e.g. severe progressive headaches for an intracranial mass, spiral fracture for physical abuse, etc.). This does not necessarily deviate from the traditional meaning but rather provides additional specificity of meaning to an individual clinician within the context of an otherwise elusive diagnostic reasoning process. Our aim was to investigate effects of a cognitive intervention strategy of 'isolating red flags to enhance diagnosis (I-RED)' on diagnostic accuracy. We also secondarily aimed to assess whether I-RED strategy impacted time to decision-making and decision confidence.

Methods

Design

We conducted a 2 (I-RED strategy: Yes/No) × 2 (case complexity: simple/complex) vignette-based experiment and manipulated I-RED strategy between subjects and case complexity within subjects (Fig. 2). The I-RED strategy involved a prompt that instructed participants to isolate any red flags from clinical data before formulating a differential diagnosis. In absence of the I-RED strategy (i.e. control) condition, this instruction was omitted. The order of case complexity was counterbalanced in both experimental and control groups (see Fig. 2). Following exposure to a practice case, I-RED strategy and case complexity were combined orthogonally to create 4 groups with the following presentations: I-RED strategy for a simple case followed by a complex case, I-RED strategy for a complex case followed by a simple case, no I-RED strategy for a simple case followed by a complex case and no I-RED strategy for a complex case followed by a simple case. We used PROC POWER in SAS(R) to determine the sample size needed to detect a difference of 20% in the proportion of correct diagnoses between the I-RED strategy and control groups with a power of 0.7. A total of 154 participants was needed for the study, in case of an order effect. The study protocol was approved by the Baylor College of Medicine and Indiana University institutional review boards.

Experimental module development

The intervention consisted of the I-RED strategy administered at two levels of case complexity. We developed a computerized module designed to emulate the diagnostic process of clinicians interacting with a patient's electronic medical record (EMR). More specifically, participants could access clinical information by navigating through index tabs within the simulated EMR.

In regards to case complexity, a panel of four experienced attending physicians used an iterative process to select five pediatric clinical cases from a pool of clinical cases that represented varying levels of complexity. All five cases were developed from real cases representing 'do-not-miss diagnoses' in pediatric practice. The cases and the module were pilot tested with pediatric critical care fellows, the results (i.e. informal written comments) of which were reviewed by the expert attending physicians. After review, three cases were selected for the study: (i) a practice case (pyloric stenosis), (ii) a simple case (midgut volvulus) and (iii) a case with a complex co-morbidity (methanol toxicity in a patient with hydrocephalus). The same panel of

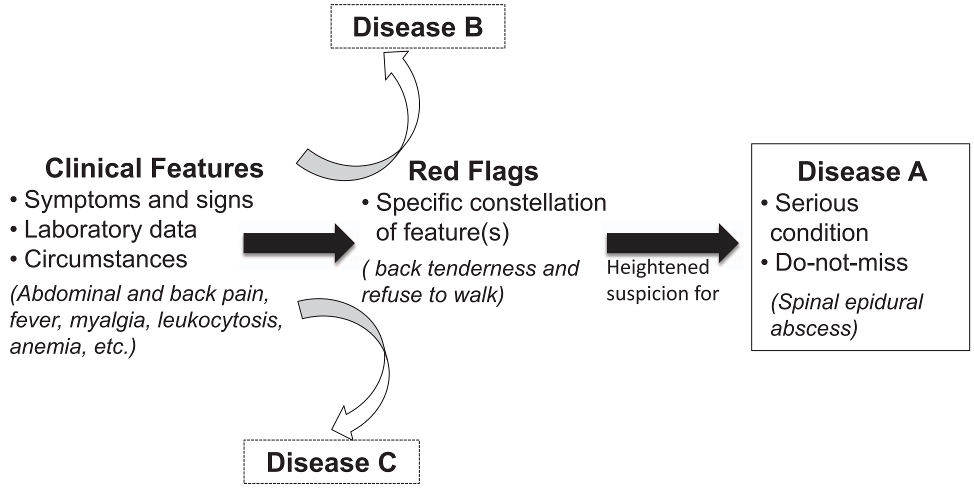

Figure 1 A conceptual framework for a red flag within a diagnostic process. A red flag is defined as a constellation of symptoms, signs, clinical data or circumstances, conceptualized by an individual clinician, that should lead to heightened suspicion for a serious condition and trigger additional evaluation.

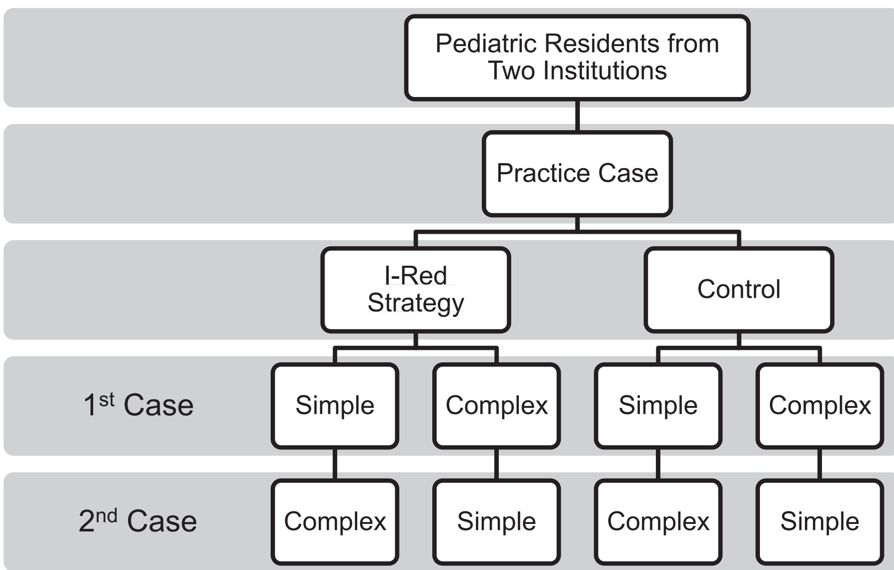

Figure 2 Study flow diagram. All participants go through the practice case (i.e. pyloric stenosis), and then are randomized to the I-RED strategy or control group, and then further randomized to either the complex case (i.e. methanol toxicity in a patient with hydrocephalus) followed by a simple case (i.e. midgut volvulus) or the simple followed by the complex case.

experienced attending physicians determined the potential red flags for each case through consensus (see vignettes in the Supplementary Material available in *INTQHC Journal* online).

Experimental procedure

We recruited pediatric residents in all three post-graduate levels of training from two university-based residency programs in the United States to participate in the study between October 2015 and December 2016. Residents involved in the development of the computerized module or the cases were excluded. The participants were randomly assigned using a computer-generated random number assigner.

All participants completed a practice case to become familiar with the procedural aspects of the study. We purposefully used a routine scenario for the practice case to prevent a Hawthorne effect to the participants (i.e. being alert for do-not-miss diagnoses). Subsequently, they completed the simple and complex cases, with the order of these two cases counterbalanced.

After an electronic informed consent was obtained, the participants in the control group navigated through a simulated EMR module containing a complete history and physical as well as appropriate laboratory and radiological findings readily available for them to gather, interpret and determine a differential diagnosis. The participants then gave a clinical impression via free text as usually done in a clinical setting. They were then asked to list their top three differential diagnoses in free text and the associated confidence level (1–10, with 10 being extremely confident) for each case.

The participants in the I-RED strategy group completed the same sequence of simulated case scenarios, but their sequence was paused with a prompt to isolate red flags, if present, prior to making a differential diagnosis. The operational definition of red flag developed by the investigators (i.e. a constellation of symptoms, signs, clinical data or circumstances that should lead to heightened suspicion for a serious condition and trigger additional evaluation) was displayed on the screen and then they were asked to list any number of red flags within each case. Next, the participants were asked to list their top three differential diagnoses and confidence level. Participants in both groups did not know the correct diagnoses for the cases until the completion of the study.

Outcome measures

The principal outcome measure was binary diagnostic accuracy, scored as 1 if any of the three differential diagnosis contained the correct diagnosis, and 0 if not. In addition, we recorded the total time the participant spent and associated confidence on each case, and henceforth refer to this as 'decision time' and 'decision confidence.'

In order to assess whether the I-RED strategy led to more red flag generation than the control condition, two investigators independently coded the free text of the control group for possible statements that met the definition of a red flag, in addition to confirming the correctness of the red flags listed in the I-RED strategy group for comparison. Then they discussed to arrive at consensus. The total number of red flags generated was used as an indicator of the effectiveness of the I-RED strategy intervention.

Statistical analyses

A logistic-regression model was used to study the effect of the I-RED strategy, complexity and order on diagnostic accuracy. We generated odds ratios (OR) and 95% confidence intervals (CI), comparing the I-RED strategy group to the control group, both as a main effect (averaged across the simple and complex cases) and as simple main effects within each level of case complexity. The OR serves as an index of the effect size. To study the effects on red flags generated and decision time, we used an ANOVA model and used the partial eta-square (η_p^2) statistic to describe the effect sizes.

Given that order effects are a concern in within-subjects' designs, we planned a two-part analysis. First, we limit the analysis to the first case to allow for an order-uncontaminated, between-subjects test of the focal hypothesis, i.e. the effect of I-RED strategy on diagnostic accuracy and decision confidence for simple/complex cases. We refer to this as the 'Between-Subjects Analysis.' Following this, we analyzed the results for the complete design, which includes the second case. Since the response to the first and second case came from the same participant, we used a Generalized Estimating Equations approach to account for potential within-subject correlation. We refer to this as the 'Within-Subjects Analysis.' All analyses were conducted using SAS® Version 9.4.

Results

Of 195 eligible residents, 118 residents signed up, and 109 (65 and 44 from institution one and two, respectively) completed the module. There were 56 and 53 participants in the control and I-RED group, respectively, each providing 2 judgments, resulting in a total of 218 judgments. Assuming no order effect, this would be enough to conduct a sufficiently powered study. We first specified a model predicting diagnostic accuracy with I-RED Strategy, Case Complexity, and Order, and all their interactions as predictors. However, we observed a significant three-way effect of I-RED Strategy × Order × Case Complexity (F (1, 105) = 6.76, $P < 0.01$), suggesting that order was moderating the effect of I-RED Strategy and Case Complexity. For this reason, we limit the analyses to the order-uncontaminated, between-subjects design by using only the first case seen by each participant (not counting the practice scenario, common to all participants). The results of the order-contaminated within-subjects analyses are detailed in the Supplementary Material available in *INTQHC Journal* online.

Between-subjects analysis

Isolation of red flags. To ensure that the I-RED intervention worked as intended, we examined the effect of the I-RED strategy, case complexity and their 2-way interaction on the number of red flags generated. As expected, the I-RED strategy led to a significant increase in the number of red flags that were generated versus the control condition (3.04 vs. 2.09); F (1, 105) = 23.57, $P < 0.001$, effect size (η_p^2) = 0.18. Examples of red flags generated were 'visual disturbances' and 'high-gap metabolic acidosis' for methanol poisoning and 'bloody stool in a neonate' and 'rigid abdomen' for midgut volvulus. Case complexity did not significantly impact number of red flags generated. The interaction between the I-RED strategy X case complexity was also not significant, i.e. I-RED strategy increased the number of red flags generated regardless of case complexity.

Diagnostic accuracy. Case complexity had a significant effect on accuracy, F (1, 105) = 29.29, $P < 0.001$; the more complex case resulted in lower diagnostic accuracy compared to the simple case, (32 vs. 83% respectively; OR = 0.07, 95% CI [0.03–0.18], $P < 0.001$). However, the I-RED strategy did not statistically affect accuracy, with 62 and 48% accuracy in the I-RED versus no I-RED groups, respectively; F (1, 105) = 2.14, $P = 0.14$; OR = 2.07 [0.78–5.5]. The interaction between I-RED strategy and case complexity was also not significant, χ^2 (1 df) = 0.06, $P = 0.8$ (Table 1).

Decision time. With regard to time to complete the case, we observed a significant effect of the I-RED strategy, F (1, 103) = 10.27, $P < 0.01$, $\eta_p^2 = 0.09$; the participants in the I-RED strategy group took significantly less time than those in the control group to arrive at a differential diagnosis, 462 versus 658 s. Conversely, neither the main effect of case complexity, F (1, 104) = 2.11, $P = 0.15$, $\eta_p^2 = 0.02$, nor the I-RED strategy × case complexity interaction on time to complete the case were significant, F (1, 104) < 1, $P = 0.33$, $\eta_p^2 = 0.01$ (Table 1).

Decision confidence. We observed a significant effect of the I-RED strategy × case complexity interaction on confidence, F (1, 105) = 4.18, $P < 0.05$, $\eta_p^2 = 0.04$. For the simple case, we observed a significant difference on mean confidence between the I-RED strategy and control groups, 7.00 (standard deviation [SD] = 2.04) versus 5.77 (SD = 1.75), respectively, $P = 0.02$. In contrast, there was no difference in mean confidence between I-RED strategy and control groups for the complex case, 6.22 versus 6.47, respectively, $P = 0.62$. We did not observe a significant main effect of either I-RED strategy or case complexity.

Discussion

We conducted an experimental vignette study to determine whether a cognitive intervention, a prompt to isolate red flags to enhance diagnosis (I-RED), can improve diagnostic accuracy. Participants who were prompted to isolate red flags generated more red flags, took less time to derive a differential diagnosis and reported more confidence

Table 1. Means for the outcome measures the I-RED strategy and the control

	Control ($n = 56$)	iRED strategy ($n = 53$)	P-value
Red flags isolated (mean, SD)	2.09 (1.01)	3.04 (1.04)	$P < 0.01$
Accuracy (% correct)	48%	62%	$P < 0.14$
Time* (s, mean, SD)	658 (381)	462 (231)	$P < 0.0001$

*Time refers to time to complete the case.

in their diagnostic decision for the simple case but did not achieve statistically higher diagnostic accuracy.

Strategies for improving reasoning could reduce cognitive error and in turn reduce diagnostic errors [17–24]. Mamede et al. [20] reported an application of diagnostic reflection that improved diagnoses in an experimental setting. Use of checklists, evidence-based diagnostic protocols and algorithms in a standardized, consensus-based fashion has been suggested [25, 26]. Most decision support tools and checklists are underused and rarely incorporated into the workflow [27, 28]. Cognitive strategies to improving analytical reasoning through prompts to isolate red flags to enhance differential diagnoses have not been evaluated [9]. Red flag isolation for 'do not miss' diagnoses could be useful in complex cases where cognitive overload can lead to flawed reasoning in the diagnostic process [29]. Our findings suggest that this intervention is worthy of additional development and evaluation despite not achieving statistical significance, especially given potential positive effects on efficiency and confidence.

Decision time was significantly shorter with the I-RED strategy, suggesting improved efficiency, but it is unclear how this translates to the real world. Both accuracy and efficiency are of importance in the increasingly time-pressured patient–physician encounter. The important role of efficiency has been underscored in the literature both empirically [30, 31] and conceptually [32]. This is also echoed in the NASEM report, which explicitly states that improving diagnostic performance will require addressing both diagnostic quality and efficiency in order to achieve high-value diagnostic performance [1]. Further investigation is needed to explore the impact of shortened decision time with I-RED strategy and how to achieve an optimal balance between efficiency and accuracy.

Our negative findings on diagnostic accuracy should be seen as inconclusive rather than definitive because the study may have been under-powered to find a clinically important difference. First, we planned our study with an expected power of 0.7 but obtained a smaller than needed sample size. Also, the vignettes and certain circumstantial factors in the construction and delivery of the clinical vignette may have contributed to variability. Specifically, we aimed to replicate the real-world clinical experience of diagnostic decision-making to improve validity. We hosted the experiment on an online platform rather than in a controlled laboratory-like setting. This allowed the study to be presented within the context of an EMR, at least somewhat similar to what participants might experience in real-world clinical setting. We observed wide variability in the amount of time participants took to solve each vignette, and some solved the clinical vignettes in more than one sitting. These likely added to unintended variability in how the experiment was experienced by the participant, as indexed by the time to completion. Another limitation is that our participants were in different levels of training, though this should not skew our results due to randomization of the participants. As participants completed the experiment remotely, we were also not able to confirm whether the participant used additional resources or asked others for help to develop the differential diagnoses. The data collection took over a year so there might have been opportunities that participants informed one another about the cases.

Despite these limitations, our study yielded insights for future research on interventions to improve diagnostic decision-making. First, the inherently complex nature of diagnostic decision-making as well as the multiple circumstantial factors around this process pose significant challenges to conducting a sufficiently controlled experiment. This presents a dilemma for the researchers who have to choose between excluding environmental factors in simulated settings versus studying decision-making in its natural habitat. Investigating an intervention in an experimental environment could nevertheless be an important step before testing or implementing it in clinical settings. While the latter makes research messier, it is important to embrace the complexity of the diagnostic process both in modeling the psychological processes (e.g. recognize, hypothesize and test for order effects) and in studying procedures by situating them in realistic clinical settings. This raises the issue of study power, which should ideally be addressed by increasing the sample size.

Conclusion

A cognitive strategy of isolating red flags prior to differential diagnosis did not improve diagnostic accuracy related to 'do-not-miss diagnoses' in this experimental vignette study. However, the strategy led to the generation of more red flags and higher reported confidence in diagnostic decisions for the simple case. Given the paucity of evidence-based solutions to reduce diagnostic error, findings warrant additional exploration of this intervention.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

References

1. National Academies of Sciences Engineering and Medicine, *Improving Diagnosis in Health Care*, 2015.
2. Zwaan L, Bruijne MC, Wagner C *et al.* A record review on incidence, consequences and causes of diagnostic adverse events. *Arch Intern Med* 2010;**170**:1015–21.
3. Singh H *et al.* Errors of diagnosis in pediatric practice: a multisite survey. *Pediatrics* 2010;**126**:70–9.
4. Singh H *et al.* Identifying diagnostic errors in primary care using an electronic screening algorithm. *Arch Intern Med* 2007;**167**:302–8.
5. Singh H, Giardina TD, Forjuoh SN *et al.* Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ Qual Saf* 2012;**21**:93–100.
6. 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.
7. McDonald KM *et al.* Patient safety strategies targeted at diagnostic errors: a systematic review. *Ann Intern Med* 2013;**158**:381–9.
8. Graber ML *et al.* Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012;**21**:535–57.
9. Pelaccia T *et al.* An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online* 2011;16:10.3402/meo.v16i0.5890.
10. Thammasitboon S, Thammasitboon S, Singhal G. Diagnosing diagnostic error. *Curr Probl Pediatr Adolesc Health Care* 2013;**43**:227–31.

11. Thammasitboon S, Thammasitboon S, Singhal G. System-related factors contributing to diagnostic errors. *Curr Probl Pediatr Adolesc Health Care* 2013;**43**:242–7.
12. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med* 2011;**86**:307–13.
13. Graber ML. Educational strategies to reduce diagnostic error: can you teach this stuff? *Adv Health Sci Educ Theory Pract* 2009;**14**: 63–9.
14. Meyer AN *et al*. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med* 2013;**173**:1952–8.
15. Bhise V *et al*. Errors in diagnosis of spinal epidural abscesses in the era of electronic health records. *Am J Med* 2017;**130**:975–81.
16. Singh H *et al*. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013;**173**:418–25.
17. Croskerry P. *Diagnostic Failure: A Cognitive and Affective Approach*. AHRQ, 2005.
18. Croskerry P. The cognitive imperative: thinking about how we think. *Acad Emerg Med* 2000;**7**:1223–31.
19. Moulton CA *et al*. Slowing down when you should: a new model of expert judgment. *Acad Med* 2007;**82**:S109–16.
20. Mamede S *et al*. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;**304**:1198–203.
21. Schön DA. *Educating the Reflective Practitioner*. San Francisco, Jossey-Bass, 1987.
22. Mamede S, Schmidt HG, Rikers R. Diagnostic errors and reflective practice in medicine. *J Eval Clin Pract* 2007;**13**:138–45.
23. Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. *Qual Saf Health Care* 2006;**15**:159–64.
24. Hamm RM, Zubialde J. Physicians' expert cognition and the problem of cognitive biases. *Prim Care* 1995;**22**:181–212.
25. Zwaan L *et al*. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Acad Med* 2012;**87**:149–56.
26. Ely JW, Graber MA. Checklists to prevent diagnostic errors: a pilot randomized controlled trial. *Diagnosi* 2015;**2**:163–9.
27. Miller RA. Computer-assisted diagnostic decision support: history, challenges, and possible paths forward. *Adv Health Sci Educ Theory Pract* 2009;**14**:89–106.
28. Rosenbloom ST *et al*. Effect of CPOE user interface design on user-initiated access to educational and patient information during clinical care. *J Am Med Inform Assoc* 2005;**12**:458–73.
29. Croskerry P, Sinclair D. Emergency medicine: a practice prone to error? *CJEM* 2001;**3**:271–6.
30. Sherbino J *et al*. The relationship between response time and diagnostic accuracy. *Acad Med* 2012;**87**:785–91.
31. Norman G *et al*. The etiology of diagnostic errors: a controlled trial of system 1 versus system 2 reasoning. *Acad Med* 2014;**89**:277–84.
32. Newman-Toker DE, McDonald KM, Meltzer DO. How much diagnostic safety can we afford, and how should we decide? A health economics perspective. *BMJ Qual Saf* 2013;**22**:ii11–20.