BMJ Open

Checklists to reduce diagnostic error: a systematic review of the literature using a human factors framework

Jawad Al-Khafaji [iD] ,[1,2] Ryan F Townshend,[3] Whitney Townsend,[4] Vineet Chopra,[5] Ashwin Gupta[1,2]

[1]Department of Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA
[2]Department of Medicine, Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, Michigan, USA
[3]University of Michigan Medical School, Ann Arbor, Michigan, USA
[4]Taubman Health Sciences Library, University of Michigan, Ann Arbor, Michigan, USA
[5]Department of Medicine, University of Colorado School of Medicine, Aurora, Colorado, USA

Correspondence to
Jawad Al-Khafaji;
jawadal@med.umich.edu

ABSTRACT

Objectives To apply a human factors framework to understand whether checklists to reduce clinical diagnostic error have (1) gaps in composition; and (2) components that may be more likely to reduce errors.

Design Systematic review.

Data sources PubMed, EMBASE, Scopus and Web of Science were searched through 15 February 2022.

Eligibility criteria Any article that included a clinical checklist aimed at improving the diagnostic process. Checklists were defined as any structured guide intended to elicit additional thinking regarding diagnosis.

Data extraction and synthesis Two authors independently reviewed and selected articles based on eligibility criteria. Each extracted unique checklist was independently characterised according to the well-established human factors framework: Systems Engineering Initiative for Patient Safety 2.0 (SEIPS 2.0). If reported, checklist efficacy in reducing diagnostic error (eg, diagnostic accuracy, number of errors or any patient-related outcomes) was outlined. Risk of study bias was independently evaluated using standardised quality assessment tools in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Results A total of 30 articles containing 25 unique checklists were included. Checklists were characterised within the SEIPS 2.0 framework as follows: Work Systems subcomponents of Tasks (n=13), Persons (n=2) and Internal Environment (n=3); Processes subcomponents of Cognitive (n=20) and Social and Behavioural (n=2); and Outcomes subcomponents of Professional (n=2). Other subcomponents, such as External Environment or Patient outcomes, were not addressed. Fourteen checklists examined effect on diagnostic outcomes: seven demonstrated improvement, six were without improvement and one demonstrated mixed results. Importantly, Tasks-oriented studies more often demonstrated error reduction (n=5/7) than those addressing the Cognitive process (n=4/10).

Conclusions Most diagnostic checklists incorporated few human factors components. Checklists addressing the SEIPS 2.0 Tasks subcomponent were more often associated with a reduction in diagnostic errors. Studies examining less explored subcomponents and emphasis on Tasks, rather than the Cognitive subcomponents, may be warranted to prevent diagnostic errors.

Strengths and limitations of this study

⇒ This is the first review to use a human factors framework to study checklists aimed at reducing diagnostic error.
⇒ The search's broadness and inclusiveness help in finding trends among the checklists literature that's still in its infancy.
⇒ Despite its broadness and inclusiveness, elements outside Systems Engineering Initiative for Patient Safety 2.0 (SEIPS 2.0) may also be important in evaluating diagnostic checklists.
⇒ Characterising clinical diagnosis checklists according to SEIPS 2.0 can be subjective.

INTRODUCTION

Diagnostic error is a leading cause of mortality in the USA.[1–3] As a preventable source of patient harm, errors in diagnosis represent a key opportunity to improve patient safety.[1–4] Diagnostic errors are multifactorial in aetiology—individual aspects such as faulty reasoning and cognitive biases as well as system processes related to data retrieval and distractions are all posited to be associated with errors.[5–7] To reduce patient harm, several system (eg, trigger alerts in times at high risk) and individual interventions (eg, training against common biases) have been tested.[8–10] Yet, to date, successful interventions to reduce diagnostic error remain elusive.

Beyond diagnostic errors, one approach shown to be efficacious in preventing errors and improving safety is the checklist.[11] Used in myriad fields including military, agriculture and aviation,[12–15] checklists organise processes into sequential, step-by-step tasks, thus simplifying complex work. In similar fashion, human factors, a discipline established in most safety critical industries, uses knowledge about human behaviour to design safer systems.[16] In medicine, integration of checklists and human factors engineering

have been shown to improve safety in multiple domains including central line-associated bloodstream infections and surgical time-outs.[17 18] However, whether these tools affect diagnostic outcomes remains unclear.[19]

While several checklists to reduce diagnostic errors have been published,[20–22] whether and how human and systems factors are integrated within current checklist contents remains unknown. Challenges within cognitive, systems and patient factors have been identified as important root causes for diagnostic error.[23–26] One may hypothesise that checklists may reduce diagnostic error by targeting all of these root causes; conversely those that focus on one or none of these issues are less likely to be successful. Alternatively, a single component of these checklists may represent the 'active ingredient', and therefore, may be a high-yield target for future research. To date, this empiric question has not been answered. Therefore, in this systematic review, we examined checklists aimed at decreasing diagnostic errors by utilising an established human factors framework: the Systems Engineering Initiative for Patient Safety 2.0 (SEIPS 2.0).[27] The SEIPS 2.0 model has been widely adopted in healthcare research to advance patient safety and quality improvement.[28–30] We used SEIPS 2.0 to study available clinical diagnosis checklists and understand which human factors are more associated with reducing diagnostic error.

METHODS

We developed a review protocol (PROSPERO: CRD42019136830) and followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses recommendations for reporting our findings.[31]

Data sources and searches

A medical librarian (WT) performed serial literature searches between 30 April 2019 and 15 February 2022 for articles containing clinical diagnosis checklists. Databases were searched from inception and included PubMed, EMBASE, Scopus and Web of Science. Searches were designed for each database and included controlled vocabulary terms (eg, Medical Subject Headings), keywords to represent concepts, including "checklist," "differential diagnosis" and "diagnostic errors." No restrictions were placed on publication date, language or completion status. The full search strategy and code are available in the attached online supplemental file 1.

Study selection

Articles were eligible for inclusion if they included one or more checklists aimed at improving the diagnostic process. To ensure rigour, in addition to articles where checklists were formally evaluated via experimental designs, we also included articles that proposed checklists without formal evaluation. Articles were excluded if they did not have an explicit checklist or if their checklists were solely for disease screening purposes; included checklists that were only intended to diagnose a specific

disease (eg, melanoma); or if they focused on processes outside of diagnosis (eg, procedural/surgical safety checklists). If the same checklist was used in more than one study, each individual study was included provided that the study design and methods reporting checklist performance were substantively different. Checklists were defined broadly and included any multicomponent structured guide intended to elicit additional thinking regarding diagnosis. This included disease/symptom specific differential diagnosis checklists that incorporated a comprehensive list of items aimed at avoiding missing alternative diagnoses as well as general debiasing checklists aimed at avoiding missing basic steps in the diagnostic process (ie, checklists that lead to reproducible approach to diagnosis). Checklists did not necessitate physical box checking for inclusion, as only five published checklists included this requirement (online supplemental Table S1).

We considered peer-reviewed articles published in any language, but only included foreign language articles when an English translation was available. Two authors (JA-K and RFT) independently determined study eligibility; when necessary, differences were adjudicated by a third author (AG). Interrater agreement for study eligibility and data abstraction was assessed using the Cohen k coefficient.

Data extraction and quality assessment

Data were extracted from included articles independently and in duplicate by two authors (J.A. and RFT) using a template adopted from the Cochrane Collaboration[32] (online supplemental file 1). Data on the study design, population, checklist content and study outcomes were extracted. Each study, when appropriate, was assessed for risk of bias (RoB) in its design, data synthesis and analysis, outcome-measuring and conclusion, by two authors using the following established and validated quality assessment tools: Cochrane's RoB 2.0 tool for randomised studies,[33] ROBINS-I tool for non-randomised quasi experimental studies,[34] the Newcastle-Ottawa Scale for cohort studies (with control),[35] and the National Institute of Health's (NIH) Quality Assessment tool for pretest and post-test cohort studies without control (see online supplemental file 1 for references to those tools).[36] Discrepancies in ratings were resolved after joint review and discussion.

Data synthesis and analysis
Checklist categorisation

We categorized elements targeted in diagnostic checklists using the SEIPS 2.0 framework. We chose SEIPS 2.0 as an organising structure to assess checklists because it considers clinical aspects, cognitive, human and system factors that can contribute to diagnostic errors and adverse events.[37] The three primary components of SEIPS 2.0 include: Work System (including subcomponents of Person(s), Tasks, Internal Environment, External Environment, Tools and Technologies and Organisation), Processes (including subcomponents of Cognitive,

Physical, and Social and Behavioural) and Outcomes (including subcomponents of Patient, Professional and Organisational).[27]

We examined included elements of each study checklist and assigned them to components or subcomponents of SEIPS 2.0. Within the Work System, checklist elements were characterised as Persons if they address factors such as clinicians' level of clinical knowledge, clinicians or patients' demeaner, or patient factors such as health status, age and level of education. Elements requesting that physicians perform actions to assist in the diagnostic process (eg, obtain history, order a test) were classified as Tasks. Checklist elements that focused on care setting (eg, location of care, workspace design and noise) were considered as targeting the Internal Environment, while those elements focused on the larger or the non-immediate environment (eg, leadership decisions and institutional policies) were categorised as targeting the External Environment. Elements advocating the use of tools (eg, digital aids) to assist in the diagnostic process were assigned to the Tools and Technology category. Finally, checklists that addressed factors, such as roles of providers, patient visit times or workload, were characterised as Organisation factors.

Within the Processes component, checklist elements targeting cognitive functions (eg, listing differential diagnosis, pausing to consider cognitive biases) were categorised as Cognitive Processes, whereas elements incorporating physical factors to aid diagnosis (eg, having to a computer to access patient charts readily) were categorised as Physical Processes. Elements advocating clinician communication with either team members or patients were categorised as Social and Behavioural subcomponents. Finally, within the Outcomes SEIPS 2.0 component, checklist elements were described as incorporating Patient (eg, patient satisfaction and quality of care), Professional (eg, clinicians' fatigue and burnout), or Organisational (eg, employee turnover, staffing difficulties and compliance with regulations) Outcomes.

As most checklists on diagnostic errors focus on the cognitive contributions to error,[20] we further described the cognitive components of included checklists as follows: (1) those that posed a list of differential diagnoses; (2) those that focused on urgent medical conditions; (3) those that included history, physical exam or tests and (4) those that highlighted common diagnostic pitfalls. These categories were created following a prospective review of selected checklists for factors known to reduce diagnostic error. For example, recognising diagnostic pitfalls and systematically narrowing down differential diagnoses are cognitive strategies to reduce diagnostic error.[38 39] Given substantial heterogeneity within the included studies, formal metanalysis was not performed.

Checklist effectiveness in reducing diagnostic error

Articles that evaluated and reported outcomes (eg, diagnostic accuracy, number of diagnostic errors and any patient-related outcomes) on checklists effectiveness in reducing diagnostic error were categorised according to their study design and their outcomes. Those checklists were further classified based on inclusion of SEIPS 2.0 components for the purpose of identifying checklist components that may be associated with success or failure in reducing error.

Patient and public involvement

No patient was involved in this study.

RESULTS

Search results and study details

The serial searches yielded 5761 citations, of which 2575 were duplicates and were removed. After title and abstract screening, 69 articles underwent full-text review (1 of which)[40] was found in a systematic review.[41] Of these 69 articles, a total of 30 met all inclusion criteria. Five articles included more than one checklist. Ten articles included previously published checklists, but in each article, the checklists were evaluated using different study designs. These studies were therefore retained. After excluding duplicate checklists, a total of 25 unique checklists described within 30 articles were examined (figure 1). Within the 30 included articles, 18 formally evaluated and reported outcomes related to checklist use whereas 12 articles featured checklists that were not formally evaluated. Inter-rater agreement for study eligibility for full text review was high (kappa=0.97).

Included checklists covered each of the three major framework components of the SEIPS 2.0 model: Work System (n=15 checklists), Processes (n=20) and Outcomes (n=2). Many checklists targeted one (n=13) or two (n=10) framework components; only two checklists targeted all three SEIPS 2.0 components (table 1). Checklists are listed with examples of how their components fit into SEIPS categories in online supplemental Table S2.

Checklists targeting work systems

Fifteen out of the 25 checklists addressed Work Systems.[20 22 42–53] Two of those checklists addressed the Persons subcomponent,[45 46] 13 addressed Tasks,[20 22 42 43 45 47–53] and 3 addressed Internal Environment (table 1).[44–46] Notably, no checklists focused on Tools and Technologies, or External Environment. Only one checklist examined factors within Organisation.[45] Checklists elements incorporating the Persons subcomponent typically addressed clinicians or patients' demeanour or attitude during a clinical encounter. For example, Ely's checklist for mental pause addressed factors such as physician being 'angry' or patient being 'hostile'.[45] whereas Graber's checklist prompted the clinician to consider, 'is this a patient I don't like?'.[46] Tasks were highlighted in Ely's general debiasing checklist by prompting clinicians to 'obtain…medical history' and 'perform a…physical exam'.[20] Other examples where Tasks are emphasised include Bahrami's radiology checklist (eg, asking providers to systematically assess different brains
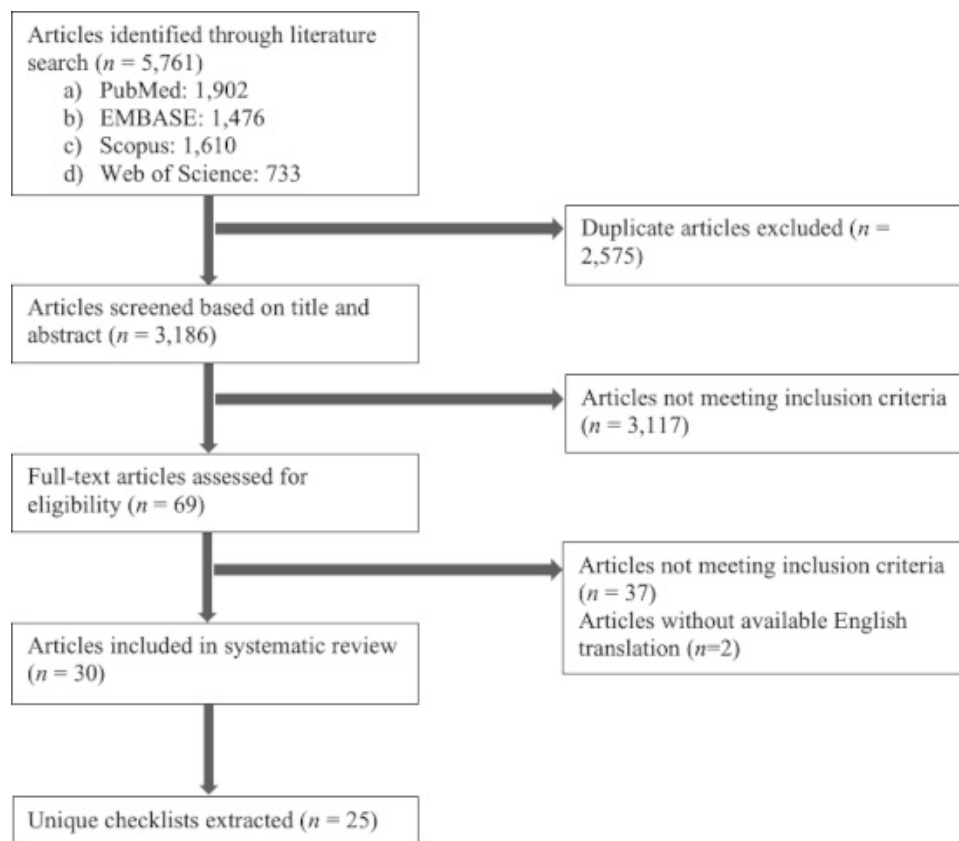
Figure 1 Study flow diagram.

areas such as sulci and ventricles),[42] Nedorost's dermatitis checklist (eg, examine for signs of dermatomyositis),[48] and Sibbald's ECG checklist (eg, calculate the rate, check the intervals PR, QRS, QT).[51] In contrast, Internal Environment was prominent in Chew's TWED (Threat, Wrong/ What else, Evidence, Dispositional factors) mnemonic (Dispositional factors, such as chaotic, busy work place).[44] Finally, the Organisation subcomponent was emphasised in Ely's general checklist, which asked the clinician about 'external pressures' (eg, time pressure), highlighting the impact of workload on clinical decision making.[45]

Checklists targeting processes framework

Most checklists (n=20) focused on the Cognitive subcomponent of Processes. No checklists addressed the Physical subcomponent and only two checklists addressed the Social and Behavioural subcomponent of Processes (table 1).[48 50]

Within the Cognitive subcomponent, checklists targeted various cognitive functions. Ten checklists included a list of differential diagnoses.[20 22 53–60] For example, Hess's lower extremity ulcer checklist linked broad areas such as 'inflammatory disorders' with related diagnoses: 'granuloma annulare, necrobiosis lipoidica and pyoderma gangrenosum.'[54] Kok's checklist focused on assisting in radiography interpretation by listing 'commonly missed diagnoses' with images including 'pneumothorax,' 'spinal cord compression' and 'surgical clips'.[22]

Six checklists emphasised broadening the history, physical exam or tests to aid in the diagnostic process.[20 22 49 53 56 58]

For example, Nordick's Diagnostic and Reasoning Tool checklist contained an annotated flow chart that walks providers through 'chief complaint,' 'history of present illness' and 'physical exam' while providing broad diagnostic questions to consider while performing these steps.[49] Similarly, Weber's checklist for orbital or periorbital cellulitis outlined specific points in the history and physical (eg, 'onset' 'catastrophic (hours))' and followed by diagnoses in order of likelihood.[58]

Three checklists specifically pointed out commonly missed diagnoses or diagnostic pitfalls.[20 22 61] For example, Ely's disease-specific differential diagnosis checklist emphasised 'commonly missed diagnosis' with an asterisk.[20] A disease-specific cognitive forcing checklist from the same paper emphasised common diagnostic pitfalls, such as missed peroneal tendon tear' and 'underappreciated ankle instability.'

Three checklists prompted clinicians to consider urgent or emergent conditions.[20 44 57] For example, the TWED mnemonic from Chew 2016 includes 'life-or-limb Threat' as a cognitive forcing tool to rule out 'worst-case scenario's.[44] Additionally, the condition-specific differential diagnosis checklist form Shimizu also highlighted 'do-not-miss-diagnosis' with a symbol.[57]

Checklists that addressed the Social and Behavioural subcomponents encouraged person-to-person communication to ensure adequate knowledge transfer. For example, Rush's CARE mnemonic incorporated 'C' to

Table 1 Categorisation of checklists components based on SEIPS 2.0*

Study ID	Study design	Checklist	Work systems	Processes	Outcomes
Bahrami 2009[42]	Expert opinion	Radiology interpretation checklist for brain	Tasks		
Bello 2019[43]	Expert opinion	Radiology interpretation checklist for skull base	Tasks		
Chew 2016[44]	Quasi-experimental	Mnemonic tool (TWED) meant to facilitate metacognition	Internal environment	Cognitive	Professional
Chew 2017[66]	Quasi-experimental	Mnemonic tool (TWED) meant to facilitate metacognition (Chew 2016)	Internal environment	Cognitive	Professional
Chew 2019[21]	Focus groups	Mnemonic tool (TWED) meant to facilitate metacognition (Chew 2016)	Internal environment	Cognitive	Professional
Ely 2011[20]	Expert opinion	General debiasing checklist	Tasks	Cognitive	
		Ely's differential diagnosis (DDx) checklists		Cognitive	
		Disease-specific cognitive forcing checklist	Tasks	Cognitive	
Ely 2015[61]	RCT	Ely's DDx checklists (Ely 2011)		Cognitive	
Ely 2016[45]	Expert opinion	General checklist for mental pause	Internal environment, Tasks, Organisation, Persons	Cognitive	Professional
		Ely's DDx checklists (Ely 2011)		Cognitive	
Graber 2014[46]	Pretest and post-test (interviews/user perception)	Checklist for high-risk diagnostic error	Internal environment, Persons	Cognitive	
		Ely's DDx checklists (Ely 2011)		Cognitive	
Hess 2008	Expert opinion	Lower extremity ulcer checklist		Cognitive	
Huang 2017	Pretest and post-test, focus groups, chart audits	Diagnostic pause tool	Tasks	Cognitive	
Kilian 2019[62]	Pretest and post-test	Mnemonic tool (ACT) meant to elicit diagnostic reflection		Cognitive	
Kok 2017[22]	RCT	Chest radiograph interpretation checklist	Tasks	Cognitive	
Li 2022[60]	Expert opinion	Checklist of causes of abdominal pain		Cognitive	
Lv 2022[59]	Case examples	Checklist of causes of abdominal pain (Li 2022)		Cognitive	
Nedorost 2018[48]	Observational and survey	Dermatitis checklist	Tasks	Social and Behavioural	
Nickerson 2019	RCT	Electrocardiogram (ECG) syncope checklist		Cognitive	
Nordick 2020[49]	Expert opinion	Diagnostic and Reasoning Tool	Tasks	Cognitive	
O'Sullivan 2018[69]	Expert opinion	Debiasing checklist		Cognitive	

Continued

Table 1 Continued

Study ID	Study design	Checklist	Work systems	Processes	Outcomes
O'Sullivan 2019[40]	RCT	Mnemonic tool (SLOW) meant to slow down reasoning and counter bias		Cognitive	
Pan 2021[56]	Retrospective cohort	Abdominal pain checklist and algorithm		Cognitive	
Rush 2017[50]	Expert opinion	Mnemonic tool (CARE) meant to counter bias	Tasks	Cognitive, Social and Behavioural	
Shimizu 2013[57]	Pretest and post-test	General debiasing checklist (Ely 2011)	Tasks	Cognitive	
		Symptom-specific DDx checklist (similar to Ely's 2011 DDx checklists)	Tasks	Cognitive	
Sibbald 2013[51]	Pretest and post-test	ECG interpretation checklist	Tasks		
Sibbald 2013[52]	Pretest and post-test	Checklist for cardiac exam	Tasks		
Sibbald 2014[63]	Pretest and post-test	ECG interpretation checklist (Sibbald 2013–1)	Tasks		
Sibbald 2015[64]	RCT	ECG interpretation checklist (Sibbald 2013–1)	Tasks		
Sibbald 2019[65]	RCT	General debiasing checklist (Ely 2011)	Tasks	Cognitive	
		ECG interpretation checklist (Sibbald 2013–1)	Tasks		
Weber 1997[58]	Expert opinion	Checklist for orbital and periorbital swelling		Cognitive	
Yung 1983[53]	Expert opinion	Flow chart	Tasks	Cognitive	

*The three primary components of SEIPS 2.0 include: Work System (eg, Person (s), Tasks, Internal Environment, External Environment, Tools and Technologies, and Organisation), Processes (eg, Cognitive, Physical, and Social and Behavioural) and Outcomes (eg, Patient, Professional and Organisational).
ACT, Alternatives, Consequences, Traits; CARE, communicate, assess, reconsider, enact; RCT, randomised controlled trial; SEIPS, Systems Engineering Initiative for Patient Safety; SLOW, Sure, Look, Opposite, Worst; TWED, Threat, Wrong/What else, Evidence, Dispositional.

represent '<u>C</u>ommunicate with your team and patient,' while Nedorost's checklist prompted physicians to consider 'framing communication.'[48 50]

Checklists targeting outcomes

Two checklists target the Outcomes SEIPS 2.0 component,[44 45] both of which addressed the Professional outcomes subcomponent (table 1). For example, Chew's TWED mnemonic has 'D' for 'Dispositional factors' encompassing 'emotional—sleepiness, tiredness, anger.'[44] No checklists specifically address the Patient or Organisational subcomponents.

Effectiveness of checklists in reducing diagnostic error

Eighteen studies described outcomes related to the use of 14 unique clinical diagnosis checklists. Four of these studies (two pretest and post-test studies,[46 47] one survey-based study[48] and one study with a focus-group design)[21] obtained user feedback when using a checklist. The remaining fourteen, including five pretest and post-test studies,[51 52 57 62 63] six randomised controlled trials (RCTs),[22 40 55 61 64 65] two quasi-experimental studies[44 66] and one retrospective cohort study,[56] evaluated whether checklists reduced diagnostic error. The main outcomes reported within those 14 studies included diagnostic accuracy, number of errors, number of abnormalities found on imaging and patient outcomes such as length of stay (tables 2 and 3). Any study-reported statistical significance associated with those outcomes is described in table 3.

Four of the five pretest and post-test studies reported significant improvement in diagnostic accuracy with their checklists (tables 2 and 3).[51 52 57 63] For example, Sibbald's study using an ECG interpretation checklist reported a reduction in errors from 279 to 70 following implementation of the checklist (p=0.01).[51] In another study of the ECG interpretation checklist by Sibbald, an average of 1.6 mistakes on ECG interpretation were fixed with checklist use (p=0.001).[63] Sibbald's cardiac exam checklist reported 51% accuracy with checklist use vs 46% diagnostic accuracy without (p=0.04)[52]; and Shimizu's symptom-specific checklists reported 67% diagnostic accuracy with the checklist vs 60% without (p<0.05).[57] One study (Kilian's ACT mnemonic) reported no significant improvements in postchecklist diagnostic accuracy.[62]

Only two of the six RCTs included in our review reported significant improvement in diagnostic accuracy. For example, Kok's radiograph interpretation checklist assisted clinicians in finding 50.1% of abnormalities on X-rays with multiple abnormalities, compared with 41.9% when the checklist was not used (p=0.04).[22] Sibbald's ECG interpretation and diagnostic pause checklist reported that 27% of ECG interpretation errors were corrected with checklist use, vs only 4% without use (p=0.01).[64] Four of the six RCTs did not demonstrate significant reduction in diagnostic errors.[40 55 61 65]

Of the quasi-experimental studies, Chew's TWED checklist tested diagnostic accuracy using five case scenarios and reported higher scores (18.5 vs 12.5 out of 50, p=0.001) with use of the checklist.[44] However, a similar study using 10 case scenarios showed no improvement in diagnostic accuracy.[66] Finally, Pan's retrospective cohort study assessing an abdominal pain checklist demonstrated significant improvement in diagnostic accuracy with checklist use (94.8% vs 82% diagnostic accuracy, p=0.034, in one subgroup; and 95.3% vs 86%, p=0.001 in another subgroup).[56]

Within the 14 studies testing checklists, most checklists targeted the Cognitive subcomponent (n=10)[22 40 44 55–57 61 62 65 66] followed by Tasks (n=7)[22 51 52 57 63–65] (table 3). Two studies targeted the Internal Environment and the Professional Outcomes subcomponents.[44 66] None of the included checklists targeted the Persons, External Environment, Tools and Technology, Organisation, Physical or Social and Behavioural subcomponents. Though the Cognitive subcomponent was the most emphasised, only four out of the 10 studies demonstrated reduction in diagnostic error,[22 44 56 57] with 1 of those four studies showing mixed results (ie, 1 checklist demonstrating reduction in diagnostic error and another failing to reduce error).[57] Conversely, five out of the seven studies targeting Tasks demonstrated significant reduction in diagnostic error (table 3).[22 51 52 63 64] In examination of the four unique checklists included in the seven studies targeting Tasks, three checklists were found to reduce diagnostic error,[22 51 52] (though one failed to reduce diagnostic error in one study,[65] but was shown to reduce diagnostic error in several other studies),[51 63 64] and one checklist failed to reduce diagnostic error (table 3).[57]

Quality assessment

Three of the six RCTs were at 'Low Risk' of bias,[22 55 64] while the other three RCTs had 'Some Concerns'.[40 61 65] The two quasi experimental studies were found to be at 'Low Risk' of study bias (online supplemental Table S3).[44 66]

Out of the five pre–post studies, two were rated as 'good' quality,[52 63] two were found to be 'fair,'[51 57] and one was rated as 'poor' quality.[62] Finally, Pan's retrospective cohort was assessed to be 'good quality'[56] (online supplemental Table S3).

DISCUSSION

Reducing diagnostic error is of paramount importance in medicine as it is a significant, yet preventable, source of mortality. Prior reviews on interventions intended to reduce diagnostic error describe cognitive and system-related interventions, but included small numbers of published clinical diagnosis checklists in this field.[67 68] Since then, a plethora of additional studies targeting diagnostic error using checklists have been published. Despite this fact, the importance of checklists within the realm of diagnostic error remains unclear. In this study, we included 30 studies and 25 unique checklists, characterised each checklist based on SEIPS 2.0 framework, and assessed which human factors may have a role in

Table 2 Characteristics of studies evaluating effectiveness of checklists

Study ID	Checklist	Participants	Setting	Outcome
Chew 2016	Mnemonic tool (TWED) meant to facilitate metacognition	Medical Students	Experimental	Checklist group scored significantly higher on a five-case scenario test compared with control group without the checklist (18.50 vs 12.50, respectively)
Chew 2017	Mnemonic tool (TWED) meant to facilitate metacognition	Medical Students	Experimental	No significant difference with or without the checklist in a script concordance test consisting of 10 cases with three response items per case. There was only a significant difference in the checklist group score when looking at the first 5 cases compared with the group without the checklist (9.15 vs 8.18, respectively).
Chew 2019	Mnemonic tool (TWED) meant to facilitate metacognition	Medical Students and Medical Doctors	Clinical	Findings from four separate focus groups suggest that the TWED mnemonic was easy to use and effective in promoting metacognition.
Ely 2015	Ely's differential diagnosis (DDx) checklists (Ely 2011)	Primary Care Physicians	Clinical	No significant difference in diagnostic error rate between physicians using checklist and those not (11.2% and 17.8% respectively), but checklist did prompt consideration of a greater number of diagnoses per patient (6.5 with checklist vs 3.4 without).
Graber 2014	Checklist for high-risk diagnostic error	Emergency Room Physicians	Clinical	Interviews demonstrated that the majority of checklist use was to help confirm original considerations and had no major impact on the final diagnosis (only 10% of usages resulted in change to the working diagnosis). One-third of usages prompted consideration of novel diagnoses.
	Ely's DDx checklists (Ely 2011)	Emergency Room Physicians	Clinical	
Huang 2017	Diagnostic pause tool	Primary Care Physicians and Nurse Practitioners	Clinical	Diagnostic pause evoked new diagnostic actions in 13% of alerts and resulted in 13% of alerted cases showing diagnostic discrepancies at a 6 month chart audit. Participants reported good integration and minimal interruption of using tool.
Kilian 2019[62]	Mnemonic tool (ACT) meant to elicit diagnostic reflection	Emergency Medicine Residents	Experimental	Emergency medicine residence reviewing eight vignettes altered their provisional diagnosis 13% after using the ACT checklist; however, this did not demonstrate any change in diagnostic error between the provisional diagnosis and the post-checklist diagnosis.
Kok 2017[22]	Chest radiograph interpretation checklist	Medical Students	Experimental	Medical students using the checklist found more abnormalities on chest radiographs with multiple abnormalities (50.1%) compared with the group without the checklist (41.9%). Of note, there was no difference between groups in images containing no abnormalities or a single abnormality.
Nedorost 2018[48]	Dermatitis checklist	Dermatologist (Principal Investigator)	Clinical	Surveys were used to gauge clinician experience after using the checklist. 8 of 15 clinicians surveyed indicated increased efficiency of diagnostic work-up. 10 patients were shown the checklist and in 6 of these instances clinicians reported improved patient engagement. In at least 2 cases, checklist lead to definitive diagnosis on the first visit.

Table 2 Continued

Study ID	Checklist	Participants	Setting	Outcome
Nickerson 2019	Electrocardiogram (ECG) syncope checklist	Emergency Medicine Residents	Experimental	No significant difference (p=0.19) was found in overall score of residents who read the ECGs with the checklists (median score 7.2; SD 1.4) vs those who read without the checklist (median score 6.8; SD 1.6). There were some significant improvements with checklist use in post-hoc assessment of recognition of Brugada, long QT, heart block and hypertrophic obstructive cardiomyopathy (HOCM) in the ECG readings. Checklist group was more likely to overread normal ECGs as abnormal.
O'Sullivan 2019[40]	Mnemonic (SLOW) meant to slow down reasoning and counter bias	Medical Professionals	Experimental	No significant difference in error rates between checklist and non-checklist groups (2.8 and 3.1 cases correct respectively out of 10 cases total).
Pan 2021[56]	Abdominal pain checklist and algorithm	Deputy Chief Physicians	Clinical	Retrospectively, diagnostic outcomes were assessed for patients presenting to the Emergency Department (ED) with acute abdominal pain for the first time. Cases that were seen using the checklist were placed in the "processes thinking group" while all others went to the "traditional group." It was found that for hospitalised patients (emergency level 2 and 3) in the processes thinking group there was a significant improvement in diagnostic accuracy as well as patient outcomes and a reduction in length of stay and average hospital expenses.
Shimizu 2013[57]	General debiasing checklist (Ely 2011)	Medical Students	Experimental	Medical students significantly increased their average proportion of correct diagnosis in five case scenarios after using the Differential Diagnosis Checklist (67% correct) compared with initial diagnosis before any checklist (60% correct) even if they had used the Debiasing Checklist beforehand. There was no statistically significance between initial diagnosis and after use of the Debiasing Checklist (62% correct).
	Symptom-specific DDx checklist (similar to Ely's 2011 DDx checklists)	Medical Students	Experimental	
Sibbald 2013[51]	ECG interpretation checklist	Cardiology Fellows	Experimental	Checklist use resulted in a statistically significant lower error rate of 0.39 per ECG interpreted compared with 1.04 without checklist. Use also increased average interpretation and verification time (94 s vs 83 s without) but did not affect surveyed cognitive load.
Sibbald 2013[52]	Checklist for cardiac exam	Medical Students	Experimental	Statistically significant increase in diagnostic accuracy pre-checklist (46%) compared with post-checklist (51%) in the setting of examining a cardiac simulator; however, this benefit was restricted to residents that were allowed to re-examine the simulator while using the checklist.
Sibbald 2014[63]	ECG interpretation checklist (Sibbald 2013–1)	Medical Students, Internal Medicine Residents, and Cardiology Fellows	Experimental	Participants pre-checklist made an average of 2.9 errors per ECG. After using the checklist participants fixed a statistically significant mean number of 1.6 mistakes.

Table 2 Continued

Study ID	Checklist	Participants	Setting	Outcome
Sibbald 2015[64]	ECG interpretation checklist (Sibbald 2013–1)	Cardiology Residents	Experimental	Checklist use was associated with higher error correction compared with an analytic prompt (0.27 errors corrected per ECG vs 0.04 respectively) as well as greater scrutiny of key variables of the ECG as found per eye-tracking.
Sibbald 2019[65]	General checklist targeting bias (Ely 2011)	Emergency Medicine Residents, Internal Medicine Residents and Cardiology Fellows	Experimental	No significant difference of error rates between content-specific checklist, process-focused checklist and no checklist groups when interpreting 20 ECGs even when cognitive biases were incorporated into cases.
	ECG interpretation checklist (Sibbald 2013–1)	Emergency Medicine Residents, Internal Medicine Residents and Cardiology Fellows	Experimental	

ACT, Alternatives, Consequences, Traits; SLOW, Sure, Look, Opposite, Worst; TWED, Threat, Wrong/What else, Evidence, Dispositional.

influencing diagnostic errors. The cognitive subcomponent of SEIPS 2.0 was targeted most frequently (20 checklists), followed by tasks (13 checklists). Other SEIPS 2.0 subcomponents were either minimally or not at all represented. When examining efficacy, checklists targeting Tasks appeared to be most associated with reductions in diagnostic errors. Taken together, these findings highlight an opportunity to further examine the effect of task-based checklists on diagnostic error. In addition, other human factors subcomponents (eg, Persons, External Environment, Tools and Technology, Organisation, Physical, or Social and Behavioural) remain unaddressed in available checklists, suggesting that more research in these areas is necessary.

Although promising, the notion that diagnostic checklists can help improve diagnostic safety has come under scrutiny.[15] Sceptics suggest that, unlike other successful patient safety checklists that target specific tasks of execution (eg, procedure time-out checklists), diagnostic safety checklists are too broad or nonspecific when they target cognitive processes.[15 19] Furthermore, diagnostic checklists may be unable to provide sufficient content assistance to help with errors of planning, and may have potential negative effects, such as added time pressure and risk of overdiagnosis.[19] Lack of a clear link between diagnostic checklists and reduction in errors may also represent limitations with study design and quality. Indeed, when examining risk of study bias, half of the RCTs (n=3) included in this review were rated as having 'Some Concerns', due to differences in baseline characteristics or limitations in the analyses to estimate the effect of the checklist intervention. In addition, three of the five pre–post test studies were rated as either 'fair' or 'poor', due to

lack of clear selection criteria of study population, the sample being insufficient, or otherwise not representative of the general population (online supplemental Table S1). Furthermore, most checklists were tested in a non-clinical 'experimental' setting, limiting applicability to the clinical realm.

Our systematic review is an important step forward in understanding the current state of checklists and advances the science in several ways. First, the finding that task-based checklists perform better suggests promise for this approach in domains where critical sequential steps (eg, interpreting ECGs and radiographic images) are needed. Larger scale implementation of these tools using rigorous study designs is therefore important. Second, the fact that only a few studies that used cognitive checklists led to improved diagnosis argues that attempts to inform complex, multifaceted decisions via a rigid checklist may not be possible. Third, we found that multiple aspects relevant to diagnosis including Tools, Technology, and External Environment Work Systems, Physical Processes, and Patient and Organisational Outcomes were not targeted by current checklists. Future iterations of checklists in diagnostic errors may explore these areas to enhance effectiveness. For example, should efforts to curb diagnostic error at the clinician level be coupled with interventions targeting hospital leadership or the environment in which clinicians work? If so, how might checklists be helpful in those domains?

Our review has limitations. First, while the human factors framework (SEIPS 2.0) is broad and inclusive in its categorisation, elements not included in this structure may also be important in evaluating diagnostic checklists. Furthermore, components of SEIPS 2.0 have not been

Table 3 Categorisation of checklists that did and did not (*italics*) improve diagnostic error

Study ID	Study design	Checklist	Work systems	Processes	Outcomes
Chew 2016[44]	Quasi-experimental	Mnemonic tool (TWED) meant to facilitate metacognition	Internal environment	Cognitive	Professional
Kok 2017[22]	RCT	Chest radiograph interpretation checklist	Tasks	Cognitive	
Pan 2021[56]	Retrospective cohort	Abdominal pain checklist and algorithm		Cognitive	
Shimizu 2013[57]*	Pretest and post-test	Symptom-specific differential diagnosis (DDx) checklist (similar to Ely's 2011 DDx checklists)		Cognitive	
Sibbald 2013[51]	Pretest and post-test	Electrocardiogram (ECG) interpretation checklist	Tasks		
Sibbald 2013[52]	Pretest and post-test	Checklist for cardiac exam	Tasks		
Sibbald 2014[63]	Pretest and post-test	ECG interpretation checklist (Sibbald 2013)	Tasks		
Sibbald 2015[64]	RCT	ECG interpretation checklist (Sibbald 2013)	Tasks		
Chew 2017[66]	Quasi-experimental	Mnemonic tool (TWED) meant to facilitate metacognition (Chew 2016)	Internal environment	Cognitive	Professional
Ely 2015[61]	RCT	Ely's DDx checklists (Ely 2011)		Cognitive	
Kilian 2019[62]	Pretest and post-test	Mnemonic tool (ACT) meant to elicit diagnostic reflection		Cognitive	
Nickerson 2019	RCT	ECG syncope checklist		Cognitive	
O'Sullivan 2019[40]	RCT	Mnemonic tool (SLOW) meant to slow down reasoning and counter bias		Cognitive	
Shimizu 2013[57] *	Pretest and post-test	General debiasing checklist (Ely 2011)	Tasks	Cognitive	
Sibbald 2019[65]	RCT	General debiasing checklist (Ely 2011)	Tasks	Cognitive	
		ECG interpretation checklist (Sibbald 2013)	Tasks		

*Shimizu *et al* reported outcomes on two different checklists, one that demonstrated improvement in diagnostic error and one that did not.
ACT, Alternatives, Consequences, Traits; RCT, randomised controlled trial; SLOW, Sure, Look, Opposite, Worst; TWED, Threat, Wrong/What else, Evidence, Dispositional.

tested or validated in the setting of clinical diagnosis; therefore, our findings should be viewed as preliminary and hypothesis generating for future evaluations. Second, while all authors were involved in categorising the checklists' content according to SEIPS 2.0 as objectively as possible, the categorisation process was inherently subjective. Different reviewers may come up with different categorisations. Third, many of the included checklists were not formally evaluated to determine their impact on diagnosis outcomes. While inclusion of these pre-evaluative studies allows for a more comprehensive review, studies implementing these checklists are needed and may influence our findings.

The findings of our review should spur investigators interested in addressing diagnostic error to focus checklist-based research in specific ways. Specifically, greater focus on task-based checklists appears warranted. Such approach may prioritise focusing on stepwise strategies in evaluating clinical data over the differential diagnosis of such data, such as stepwise approach to reading ECGs or chest X-rays, or evaluating a rash. Additionally, broadening checklist-based research to include unaddressed SEIPS 2.0 areas, such as the Environment, and Tools and Technologies appears necessary. These types of studies may help advance the discovery of novel opportunities, reduce error, and improve effectiveness of diagnosis.

CONCLUSION

Checklists could hold great potential to reduce clinical diagnostic error; therefore, a comprehensive understanding of their landscape is crucial for further development. Organisation of checklists by the SEIPS 2.0 is highly illustrative of the current state-highlighting areas of strength of current checklists as well as blind spots that

remain poorly addressed. Generally, current published checklists emphasise very few SEIPS 2.0 elements, mainly the Task and Cognitive subcomponents, while leaving others entirely unused. Tasks-focused checklists seem to be more associated with a significant reduction in diagnostic error. Understanding the impact of incorporating less explored SEIPS 2.0 subcomponents within checklists for diagnostic error warrants evaluation. Further, studies examining the ability of task-oriented checklists to reduce diagnostic error may ultimately facilitate improved diagnosis.

ORCID iD
Jawad Al-Khafaji http://orcid.org/0000-0002-7622-2879

REFERENCES

1 Hautz WE, Kämmer JE, Hautz SC, et al. Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room. *Scand J Trauma Resusc Emerg Med* 2019;27:54.
2 In: Kohn LT, Corrigan JM, Donaldson MS. *To err is human: building a safer health system*, 2000.
3 Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ* 2016;353:i2139.
4 Gupta A, Snyder A, Kachalia A, et al. Malpractice claims related to diagnostic errors in the hospital. *BMJ Qual Saf* 2017;27. doi:10.1136/bmjqs-2017-006774. [Epub ahead of print: 09 08 2017].
5 Wellbery C. Flaws in clinical Reasoning: a common cause of diagnostic error. *Am Fam Physician* 2011;84:1042–8.
6 Lockley SW, Barger LK, Ayas NT, et al. Effects of health care provider work hours and sleep deprivation on safety and performance. *Jt Comm J Qual Patient Saf* 2007;33:7–18.
7 Singh H, Giardina TD, Meyer AND, et al. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013;173:418–25.
8 Murphy DR, Meyer AN, Sittig DF, et al. Application of electronic trigger tools to identify targets for improving diagnostic safety. *BMJ Qual Saf* 2019;28:151–9.
9 Croskerry P. Diagnostic failure: a cognitive and affective approach. In: Henriksen K, Battles JB, Marks ES, et al, eds. *Advances in patient safety: from research to implementation (volume 2: concepts and methodology). 2005*. Advances in Patient Safety, 2005.
10 Graber M. Metacognitive training to reduce diagnostic errors: ready for prime time? *Acad Med* 2003;78:781.
11 Gawande ABooks M, ed. *The checklist manifesto : how to get things right*. 1st ed., 2010: x, 209 p.
12 United States. Federal Aviation Administration. *Pilot's handbook of aeronautical knowledge*. U.S. Dept. of Transportation, Federal Aviation Administration, 2003.
13 Kim H, Räsänen K, Chae H, et al. Analysis of checklists for agricultural safety management. *Ann Agric Environ Med* 2018;25:494–9.
14 Newkirk M, Pamplin JC, Kuwamoto R, et al. Checklists change communication about key elements of patient care. *J Trauma Acute Care Surg* 2012;73:S75–82.
15 Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. *Qual Saf Health Care* 2006;15:159–64.
16 Norris B. Human factors and safe patient care. *J Nurs Manag* 2009;17:203–11.
17 Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;355:2725–32.
18 Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009;360:491–9.
19 Zwaan LSJ. *Evidence on use of clinical Reasoning checklists for diagnostic error reduction*. Rockville, MD: Agency for Healthcare Research and Quality, 2020.
20 Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med* 2011;86:307–13.
21 Chew KS, van Merrienboer JJG, Durning SJ. Perception of the usability and implementation of a metacognitive mnemonic to check cognitive errors in clinical setting. *BMC Med Educ* 2019;19:18.
22 Kok EM, Abed A, Robben SGF. Does the use of a checklist help medical students in the detection of abnormalities on a chest radiograph? *J Digit Imaging* 2017;30:726–31.
23 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
24 Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115–20.
25 Mamede S, Schmidt HG, Rikers R. Diagnostic errors and reflective practice in medicine. *J Eval Clin Pract* 2007;13:138–45.
26 Singh H, Weingart SN. Diagnostic errors in ambulatory care: dimensions and preventive strategies. *Adv Health Sci Educ Theory Pract* 2009;14 Suppl 1:57–61.
27 Holden RJ, Carayon P, Gurses AP, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics* 2013;56:1669–86.
28 Karsh B-T, Brown R. Macroergonomics and patient safety: the impact of levels on theory, measurement, analysis and intervention in patient safety research. *Appl Ergon* 2010;41:674–81.
29 Fogarty GJ, McKeon CM. Patient safety during medication administration: the influence of organizational and individual variables on unsafe work practices and medication errors. *Ergonomics* 2006;49:444–56.
30 Lacson R, Cochon L, Ip I, et al. Classifying safety events related to diagnostic imaging from a safety reporting system using a human factors framework. *J Am Coll Radiol* 2019;16:282–8.
31 Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006–12.
32 Higgins JPT TJ, Chandler J, Cumpston M, et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.1 Cochrane*, 2020.
33 Sterne JAC, Savović J, Page MJ, et al. Rob 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.
34 Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

35 Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol* 2010;25:603–5.

36 NIH National Heart LaBl. *Study quality assessment tools*.

37 Carayon P, Schoofs Hundt A, Karsh B-T, *et al*. Work system design for patient safety: the SEIPS model. *Qual Saf Health Care* 2006;15 Suppl 1:i50–8.

38 Schiff GD, Kim S, Abrams R. Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project. In: Henriksen K, Battles JB, Marks ES, *et al*, eds. *Advances in patient safety: from research to implementation (volume 2: concepts and methodology). 2005. advances in patient safety*. Advances in Patient Safety, 2005.

39 Chishti S, Jaggi KR, Saini A, *et al*. Artificial Intelligence-Based differential diagnosis: development and validation of a probabilistic model to address lack of large-scale clinical datasets. *J Med Internet Res* 2020;22:e17550.

40 O'Sullivan ED, Schofield SJ. A cognitive forcing tool to mitigate cognitive bias - a randomised control trial. *BMC Med Educ* 2019;19:12.

41 Griffith PB, Doherty C, Smeltzer SC, *et al*. Education initiatives in cognitive debiasing to improve diagnostic accuracy in student providers: a scoping review. *J Am Assoc Nurse Pract* 2020;33:862-871.

42 Bahrami S, Yim CM. Quality initiatives: blind spots at brain imaging. *Radiographics* 2009;29:1877–96.

43 Bello HR, Graves JA, Rohatgi S, *et al*. Skull Base-related lesions at routine head CT from the emergency department: pearls, pitfalls, and lessons learned. *Radiographics* 2019;39:1161–82.

44 Chew KS, Durning SJ, van Merriënboer JJ. Teaching metacognition in clinical decision-making using a novel mnemonic checklist: an exploratory study. *Singapore Med J* 2016;57:694–700.

45 Ely JW, Graber ML. Preventing diagnostic errors in primary care. *Am Fam Physician* 2016;94:426–32.

46 Graber ML, Sorensen AV, Biswas J, *et al*. Developing checklists to prevent diagnostic error in emergency room settings. *Diagnosis* 2014;1:223–31.

47 Huang GC, Kriegel G, Wheaton C, *et al*. Implementation of diagnostic pauses in the ambulatory setting. *BMJ Qual Saf* 2018;27:492–7.

48 Nedorost S. A diagnostic checklist for generalized dermatitis. *Clin Cosmet Investig Dermatol* 2018;11:545–9.

49 Nordick CL. Integrating strategies for improving diagnostic reasoning and error reduction. *J Am Assoc Nurse Pract* 2020;33:366-372.

50 Rush JL, Helms SE, Mostow EN. The care approach to reducing diagnostic errors. *Int J Dermatol* 2017;56:669–73.

51 Sibbald M, de Bruin ABH, van Merrienboer JJG. Checklists improve experts' diagnostic decisions. *Med Educ* 2013;47:301–8.

52 Sibbald M, de Bruin ABH, Cavalcanti RB, *et al*. Do you have to re-examine to reconsider your diagnosis? checklists and cardiac exam. *BMJ Qual Saf* 2013;22:333–8.

53 Yung CY. Evaluation and differential diagnosis of movement disorders: a flow chart approach. *Brain Res Bull* 1983;11:149–52.

54 Hess CT. Checklist for differential diagnosis of lower-extremity ulcers. *Adv Skin Wound Care* 2010;23:480.

55 Nickerson J, Taub ES, Shah K. A checklist manifesto: can a checklist of common diagnoses improve accuracy in ECG interpretation? *Am J Emerg Med* 2020;38:18–22.

56 Pan H-M, Li H-L, Shen Z-S, *et al*. Observation of the effectiveness of a diagnostic model for acute abdominal pain based on the etiology checklist and process thinking. *Risk Manag Healthc Policy* 2021;14:835–45.

57 Shimizu T, Matsumoto K, Tokuda Y. Effects of the use of differential diagnosis checklist and general de-biasing checklist on diagnostic performance in comparison to intuitive diagnosis. *Med Teach* 2013;35:e1218–29.

58 Weber R, Keerl R, Rochels R, *et al*. Checklists for the differential diagnosis of orbital Affections. *Orbit* 1997;16:11–24.

59 Lv C, Zhang J-N, Wang C-Y, *et al*. The cause of abdominal pain checklist in the differential diagnosis of vascular diseases primarily presenting with abdominal pain. *Asian J Surg* 2022;45:954–6.

60 Li X-R, Lv C, Guo H, *et al*. Application of a new checklist in the differential diagnosis of abdominal pain in the Department of general medicine. *Asian J Surg* 2022;45:586–7.

61 Ely JW, Graber MA. Checklists to prevent diagnostic errors: a pilot randomized controlled trial. *Diagnosis* 2015;2:163–9.

62 Kilian M, Sherbino J, Hicks C, *et al*. Understanding diagnosis through action: evaluation of a point-of-care checklist for junior emergency medical residents. *Diagnosis* 2019;6:151–6.

63 Sibbald M, De Bruin ABH, van Merrienboer JJG. Finding and fixing mistakes: do checklists work for clinicians with different levels of experience? *Adv Health Sci Educ Theory Pract* 2014;19:43–51.

64 Sibbald M, de Bruin ABH, Yu E, *et al*. Why verifying diagnostic decisions with a checklist can help: insights from eye tracking. *Adv Health Sci Educ Theory Pract* 2015;20:1053–60.

65 Sibbald M, Sherbino J, Ilgen JS, *et al*. Debiasing versus knowledge retrieval checklists to reduce diagnostic error in ECG interpretation. *Adv Health Sci Educ Theory Pract* 2019;24:427–40.

66 Chew KS, van Merrienboer JJG, Durning SJ. Investing in the use of a checklist during differential diagnoses consideration: what's the trade-off? *BMC Medical Education. 2017/11/29* 2017;17:234.

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

68 Singh H, Graber ML, Kissam SM, *et al*. System-Related interventions to reduce diagnostic errors: a narrative review. *BMJ Qual Saf* 2012;21:160–70.

69 O'Sullivan ED, Schofield SJ. Cognitive bias in clinical medicine. *J R Coll Physicians Edinb* 2018;48:225–32.