

Debiasing and Educational Interventions in Medical Diagnosis: A Systematic Review

Arthur Tung^{1*}

Michael Melchiorre^{2*}

¹Research and Innovation, North York General Hospital, North York, Ontario, Canada

²Department of Psychology, University of Guelph, Guelph, Ontario, Canada

*Arthur Tung and Michael Melchiorre contributed to this manuscript equally.

Correspondence can be addressed to: Arthur Tung, Research and Innovation, North York General Hospital, 4001 Leslie St, North York, ON M2K 1E1, Canada. Email: Tungarthur@yahoo.com

The authors have no conflicts of interest to disclose.

No funding was used to support this work.

Word Count: 4255

Short Running Head: Debiasing in medical diagnosis

Abstract

Background: The prevalence of cognitive bias and its contribution to diagnostic errors has been documented in recent research. Debiasing interventions or educational initiatives are key in reducing the effects and prevalence of cognitive biases, contributing to the prevention of diagnostic errors. The objectives of this review were to 1) characterize common debiasing strategies implemented to reduce diagnosis-related cognitive biases, 2) report the cognitive biases targeted, and 3) determine the effectiveness of these interventions on diagnostic accuracy. **Methods:** Searches were conducted on April 25, 2022, in MEDLINE, Embase, Healthstar, and PsycInfo. Studies were included if they presented a debiasing intervention which aimed to improve diagnostic accuracy. The Rayyan review software was used for screening. Quality assessments were conducted using the JBI Critical Appraisal Tools. Extraction, quality assessment, and analysis were recorded in Excel. **Results:** Searches resulted in 2232 studies. 17 studies were included in the final analysis. Three major debiasing interventions were identified: tool use, education of biases, and education of debiasing strategies. All intervention types reported mixed results. Common biases targeted include confirmation, availability, and search satisfying bias. **Conclusion:** While all three major debiasing interventions identified demonstrate some effectiveness in improving diagnostic accuracy, included studies reported mixed results when implemented. Furthermore, no studies examined decision-making in a clinical setting, and no studies reported long-term follow-up. Future research should look to identify why some interventions demonstrate low effectiveness, the conditions which enable high effectiveness, and effectiveness in environments beyond vignettes and among attending physicians.

PROSPERO registration number: CRD42022331128

KEY WORDS: diagnosis, debiasing, medical education, medical decision-making

Introduction

Diagnostic error has long been recognized as a significant contributor to patient harm. Missed diagnoses or incorrect diagnoses are surprisingly common, with average diagnostic error rates hovering around 13–15% under both usual practice and research-directed conditions (Newman-Toker et al., 2021; Singh, 2013). These rates generally agree with autopsy studies which report that 10–20% of cases show major diagnostic discrepancies (Graber, 2013). Diagnostic errors are not only not uncommon but also dangerous; it has been demonstrated that misdiagnosis accounted for the largest fraction of severe patient injury in the United States and was twice as likely to result in patient death or disability than other error categories (Saber Tehrani et al., 2013). Moreover, this does not account for less severe, but still consequential, effects of misdiagnosis for patients, including decreased quality of life, further disease progression, and unnecessary costs of ineffective treatments.

Previous literature focusing on cognitive psychology as it relates to medicine has found that cognitive biases and errors are major factors in generating diagnostic errors (Saposnik et al., 2016). Qualitative studies have found that cognitive errors accounted for 32% of all misdiagnoses in internal medicine settings (Schiff et al., 2009), and contributed to up to 92% of all diagnostic errors in emergency medicine settings (Okafor et al., 2016). Based on a retrospective analysis of identified and reported diagnostic discrepancies, Graber and colleagues estimated that cognitive factors contributed to 74% of misdiagnoses, averaging 4.32 cognitive contributions to error per case reviewed (320 cognitive errors in 74 cases). Due to the contribution of cognitive factors to diagnostic error and the high prevalence of diagnostic error, research, identification, and implementation of effective debiasing interventions and educational initiatives which aim to decrease common cognitive pitfalls in clinical reasoning are vital to improving diagnostic accuracy and patient outcomes.

Dual-process theory, as initially proposed by William James and further developed by Daniel Kahneman and Amos Tversky, has been adapted from the social psychology field to explain how clinician cognitive errors could arise from the use of heuristics (Klein, 2008). System 1 or type 1 processing is the mode of

reasoning best described as fast, intuitive, automatic, and heuristics based. Type 1 is also the “default” mode of thinking. Cognitive psychologists estimate most spend around 95% of their time in type 1 (Lakoff & Johnson, 1999). In contrast, type 2 processing is slow, conscious, analytical, and controlled (Kahneman, 2003).

Type 1 processing has been associated with the generation of cognitive biases when the automatic and unconscious use of heuristics is inappropriate for the situation (Norman et al., 2017). However, it must be noted that the use of type 1 processing and heuristics is usually appropriate and leads to the correct responses. Furthermore, type 2 thinking may still lead to incorrect responses if applied inappropriately (Evans & Stanovich, 2013). The cause of cognitive biases is the result of failures of both type 1 and 2 processes, type 1 for generating the error and type 2 for failing to detect and correct the error (Norman et al., 2017). Therefore, interventions which seek to minimize the effects and prevalence of cognitive biases should look to further develop the clinician’s ability to recognize unchecked use of heuristics and rely more on type 2 thinking when appropriate or teach debiasing strategies to clinicians. Type 2 thinking could also be encouraged or forced with the use of planned diagnostic slowdowns or timeouts (Yale et al., 2022). Alternatively, interventions could look to implement standardized approaches to diagnosis through the assistance of structured checklists or protocols.

In recent decades, medical educators have developed and implemented various debiasing interventions and educational initiatives within the medical curriculum and graduate medical education programs. In 2013, Croskerry and colleagues categorized various published and theorized strategies for cognitive debiasing (Croskerry et al., 2013), including the education of cognitive biases, implementation of decision-making support tools and protocols (i.e. checklists, diagnostic timeouts, and reminder systems), education of debiasing strategies (i.e. metacognition and consider-the-opposite strategy), and workplace improvements (i.e. avoidance of fatigue and sleep deprivation, and group decision strategy). Since 2013, educators have attempted to implement some of the strategies presented above in usual practice and in undergraduate or graduate medical education curriculums. However, no systematic review to our

knowledge has yet to examine the effectiveness of these recent strategies in improving diagnostic accuracy or factors supporting sound clinical decision-making.

A review of the literature which reports on the effectiveness of known interventions in reducing specific cognitive biases would help guide the development and implementation of future strategies. Given the status of the research landscape outlined above, the purpose of this systematic review was to 1) characterize and categorize the most common debiasing strategies implemented to reduce diagnosis-related cognitive biases in the literature, 2) report the types of cognitive biases targeted by each intervention, and 3) determine the effectiveness these interventions have on diagnostic accuracy.

Methods

This systematic review was conducted and reported according to the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). The review's protocol was registered with PROSPERO under the registration number CRD42022331128.

Literature Search

Searches were conducted on April 15, 2022, through four electronic databases (MEDLINE, Embase, PsycINFO, and Healthstar) using the Firefox browser. Search results were exported in Research Information Systems (RIS) format and imported to Rayyan for Systematic Reviews (Ouzzani et al., 2016). See supplementary materials for the full search strategy.

Inclusion and Exclusion Criteria

Studies were included if they 1) reported the implementation of at least one intervention or educational initiative related to diagnosis, 2) focused on reducing the prevalence or effect of cognitive biases in medical diagnosis, 3) examined diagnostic decision-making in a clinical setting or hypothetical clinical setting (i.e., patient vignettes), 4) participants in the intervention included medical students, resident physicians, attending physicians, physician assistants (PA), nurse practitioners, (NP), PA students, or NP students, and 5) were classified as an original study or conference abstract. Studies were excluded if they 1) did not explicitly ascribe at least one cognitive bias to diagnostic error, 2) were classified as a letter,

review, commentary, editorial, book, or protocol, or 3) only suggested potential debiasing strategies without implementing at least one debiasing intervention.

Two screeners independently reviewed all studies using the Rayyan software program through one round of title and abstract screening and one round of full-text review. Screeners were blinded to each other's decisions until the end of each screening round, after which a discussion was held to resolve any conflicts.

Data Extraction

Two reviewers independently extracted the included studies and imported them into a preformatted Excel worksheet. Information extracted included: author names and publication year, medical specialty, number of participants and their medical experience, intervention type and nature, cognitive bias targeted, and debiasing intervention efficacy.

Quality Assessment

Internal validity and quality of included studies were assessed with JBI's "Checklist for Randomized Control Trials" or "Checklist for Quasi-Experimental Studies (Non-Randomized Experimental Studies)" tool. Additionally, each study was given a subjective rating of poor, fair, or good to describe the raters' overall appraisal.

Results

Search Results

Database searches yielded 2232 articles, 1166 of which were identified as unique by the Rayyan software program. 1142 articles were excluded during the title and abstract screening. Inter-rater reliability (IRR, average agreement) during the abstract screening was 98.88%, with a Cohen's kappa coefficient (κ) of 0.73, indicating substantial agreement. The remaining 24 articles were retrieved and underwent full-text screening, during which 7 additional articles were excluded. IRR in this round was 86.96%, with a κ of 0.59, indicating moderate agreement. 17 articles were included in the final analysis.

See figure 1 for PRISMA article selection information.

Study Characteristics

The final analysis included a total of 1530 participants across 17 studies: 849 medical students (849/1530, 55.49%), 485 residents (465/1530, 30.39%), 30 attending physicians (30/1530, 1.96%), and 186 other healthcare professionals (186/1530, 12.16%). 11 of the studies were randomized controlled trials (RCT) (11/17, 64.71%), and 6 were quasi-experimental studies (QES) (6/17, 35.29%). Among the 17 studies, 13 examined diagnostic accuracy in a hypothetical clinical setting via case vignettes (13/17, 76.47%), 2 examined knowledge of reasoning/diagnostic skills via assessments involving patient vignettes (2/17, 11.76%), 1 examined bias recognition ability in pre-diagnosed patient vignettes (1/17, 5.88%), and 1 examined attitudes towards a conference (1/17, 5.88%). Only one of the included studies involved an actual clinical aspect (1/17, 5.88%), but did not evaluate diagnostic decision-making in an actual clinical environment.

Data Quality

IRR in quality assessment was 76.5% (13/17). 4 studies were rated as poor (4/17, 23.53%), 6 studies were rated as fair (6/17, 35.29%), and 7 studies were rated as good (7/17, 41.18%). The majority of studies rated good were successful interventions; out of 7, 5 were successful (5/7, 71.43%). The inverse was observed in studies rated poor, in which the majority were unsuccessful; out of 4, 1 was successful. (1/4, 25%).

See supplementary materials for quality assessment details (McGuinness & Higgins, 2021).

Objective 1: Debiasing Strategies Implemented

Three broad categories of debiasing intervention types were identified during the analysis of the included studies. The distribution of them across the included studies was as follows: 8 of the studies aimed to increase awareness/education of a bias (8/17, 47.06%), 7 aimed to provide education of debiasing strategies (7/17, 41.18%), and 6 used a tool such as a debiasing checklist (6/17, 35.29%). 4 of the studies' debiasing intervention utilized more than one of the three intervention types (4/17, 23.53%), but none of the studies used all three (0/17, 0%). Among the three intervention types, several subcategories were

found, including classes, workshops, checklists, metacognitive strategies, and long-term educational curricula.

Refer to Table 1 for full details of the nature of implemented interventions and their effectiveness.

Objective 2: Cognitive Biases Targeted

8 different biases were targeted across the included studies. The three most reported biases, excluding general/non-specific cognitive bias, were as follows: availability bias, targeted in 5 studies (5/17, 29.41%); search satisfying, targeted in 4 studies (4/17, 23.53%); and confirmation bias, targeted in 3 studies (3/17, 17.65%). 9 studies (9/17, 52.94%) did not target specific biases but biases in general.

Metacognition, whether by education of debiasing strategy or through the implementation of checklists designed to induce reflective thinking, was found to more commonly counteract availability bias, confirmation bias, and search satisfying. No trends were found regarding any other specific bias being commonly treated with any specific debiasing intervention.

See Table 2 for identified cognitive biases and other cognitive biases mentioned in this review (Croskerry, 2003).

Objective 3: Effectiveness of Interventions

Of the 17 studies, 8 reported an increase in diagnostic accuracy attributed to at least one debiasing intervention (8/17, 47.06%), and 5 of the studies reported no improvement in diagnostic accuracy post-intervention (5/17, 29.41%). 4 studies reported improvements in areas other than diagnostic accuracy, such as knowledge of cognitive biases and potential debiasing strategies (4/17, 23.53%). Note that studies could implement more than one type of intervention. The success rates of each of the three intervention types were as follows: among interventions aiming to increase awareness/education of bias, 6 of the 8 studies reported successful improvement of diagnostic accuracy (6/8, 75%) and 2 reported ineffectiveness (2/8, 25%); among interventions aimed at providing education of debiasing strategies, 4 of the 7 reported successful improvement of diagnostic accuracy (4/7, 57.14%), 2 reported ineffectiveness (2/7, 28.57%), and 1 study did not report effectiveness (1/7, 14.29%); and among interventions involving tool use, 4 of

the 6 studies reported successful improvement of diagnostic accuracy (4/6, 66.67%) including 1 study which reported one effective checklist and one ineffective checklist, and 2 reported ineffectiveness (2/6, 33.33%). Of the 4 studies that did not directly aim to improve diagnostic accuracy (4/17, 23.53%), 2 of the studies focused on measuring participants' knowledge of cognitive/diagnostic error (2/4, 50%), 1 study examined the information-gathering ability of participants (1/4, 25%), and one study did not have an outcome measure (1/4, 25%).

Discussion

Across 17 studies included in the final analysis, three broad categories of intervention types were identified. These categories are: 1) education or increase awareness of cognitive biases, 2) education of debiasing strategies, and 3) implementation of decision-making support tools.

Characterizing identified debiasing interventions

Education or increase awareness of cognitive biases. In 2013, Croskerry and colleagues characterized the average clinician's relationship with cognitive bias in medical decision-making at the "precontemplative level" (Croskerry et al., 2013). The precontemplative stage is denoted by general unawareness of the powerful influences of cognitive biases on decision-making despite understanding how to make correct decisions (Prochaska et al., 1992). The education of biases, therefore, provides the prerequisite knowledge needed to identify instances of cognitive biases and initiate debiasing. This type of intervention assumes that knowledge of biases will translate directly into increased proficiency in identifying those biases outside of education-driven environments. However, previous interventional studies have either failed to report the effects of intervention on diagnostic accuracy (Bond et al., 2004; Ogdie et al., 2012) or reported limited intervention effectiveness in non-clinical environments (Sherbino et al., 2011).

In the present review, 8 studies utilized education of biases as one of their implemented interventions. Most studies used conventional forms of education, namely lectures (Iyer et al., 2019; Jenkins & Youngstrom, 2016; Reilly et al., 2013; Smith & Slack, 2015), group discussion (Reilly et al., 2013),

workshops (DiNardo et al., 2016; Jenkins & Youngstrom, 2012), and reading exercises on the role of bias in diagnosis (Smith & Slack, 2015). Education of biases was also often combined with other types of interventions (i.e., debiasing checklists) during workshops and educational initiatives (Smith & Slack, 2015). Other studies provided firsthand experience of cognitive biases via vignettes designed to induce bias (Mamede et al., 2020; Plackett et al., 2020; Reilly et al., 2013). One such intervention reported by Mamede and colleagues indirectly increased physician awareness of availability bias by focusing on discriminatory elements between diseases with similar presentations (Mamede et al., 2020). Therefore, this method may demonstrate that education in the process of reasoning, like other studies of this type of intervention, still plays a secondary role to education of content knowledge in preventing bias-induced diagnostic error. Though studies which do focus on educating clinicians in the process of reasoning have similarly produced positive results (Iyer et al., 2019; Jenkins & Youngstrom, 2016; Jenkins & Youngstrom, 2012). Furthermore, the study population were residents who may have benefited most from knowledge of diseases rather than knowledge of reasoning processes.

Education of debiasing strategies. Metacognition has long been suggested as the principle debiasing strategy for tackling cognitive errors in diagnosis (Croskerry, 2003). Metacognition is a reflective approach to decision-making which examines the reasoning process used to arrive at an initial judgement (thinking about thinking) (Graber, 2003). The act of reflection allows for decoupling from the intuitive type 1 process and engaging type 2 thinking to verify initial judgements by checking for conflicting evidence and alternatives (Croskerry et al., 2013). Education of other strategies, namely group decision strategy and exposure control (i.e., limiting exposure to knowledge of another clinician's working diagnosis before forming an initial impression), may also encourage reflective reasoning, though articles investigating these interventions are limited. In addition, group decision strategy alone may succumb to diagnostic momentum if the initial diagnosis is not critically examined (O'Sullivan & Schofield, 2018). 7 studies were identified as using education of debiasing strategies in their interventions. Unsurprisingly, most studies (6/7, 85.7%) within this type of intervention utilized metacognitive principles in some

capacity (Iyer et al., 2019; Jenkins & Youngstrom, 2016; Mamede et al., 2010; Payne, 2012; Shahid et al., 2017; Sherbino et al., 2014). One study did not specify the exact debiasing strategy implemented (Jenkins & Youngstrom, 2012). One approach to implementing metacognition was “think aloud” exercises where participants explain the rationale of tests and diagnoses, allowing for reflective reasoning and engagement with type 2 processing (Shahid et al., 2017). Another popular approach was consider-the-alternative strategies, which encouraged critical examination of the differential diagnosis and seeking evidence which opposed the leading diagnosis (Jenkins & Youngstrom, 2016; Mamede et al., 2010). While these approaches should theoretically engage analytical thinking and encourage the detection of cognitive biases, it is unclear if participants would call upon these methods in actual clinical situations.

Decision-making support tools. Checklists are commonly used in other industries which are susceptible to human error and cognitive biases, namely in aviation (Gawande, 2009). Though many components of checklists in high-risk industries are obvious and familiar, the routine use of them ensures a reproducible approach across all users and could prevent easily avoided disasters (Ely et al., 2011). Surgical safety checklists are one example of how the healthcare sector has successfully adopted widespread use of checklist-based interventions to reduce human errors (Armstrong et al., 2022). In medical diagnosis, different checklists can be used in different steps of diagnosis. For example, knowledge retrieval checklists can be used to ensure information gathered on patients are unbiased, differential diagnosis checklists enforce consider-the-alternative strategies when considering the leading diagnosis, and general debiasing checklists allow for reflective thinking of the reasoning process (Ely et al., 2011). Other decision-making support tools have been developed but were not discussed in the review because they do not explicitly debias. For example, computer-based support tools like DXplain suggest possible diagnoses which fit with patient symptoms and may help complete the differential diagnosis (Barnett et al., 1987). 6 studies were found to have used checklists as one of their implemented interventions, with multiple studies implementing multiple checklists. Most studies (5/6, 83.3%) reported using a general debiasing checklist (Chew et al., 2016; DiNardo et al., 2015, 2016; Shimizu et al., 2013; Sibbald et al., 2019), which

encourages metacognition and is sometimes used in conjunction with the education of metacognitive principles (Chew et al., 2016). One study only used a differential diagnosis checklist (Myung et al., 2013). In one approach, Chew and colleagues proposed a general debiasing checklist mnemonic in an attempt to enable clinicians to conveniently apply the checklists in clinical situations, arguing that the use of a more robust checklist would be cumbersome to use outside of educational environments (Chew et al., 2016). Though the practicality of these checklists was questioned by Chew and other authors, especially in simpler clinical cases (Shimizu et al., 2013), only one study attempted to assess the usability of any checklists in a clinical setting and found that the checklist was seldomly used despite favourable feedback from clinician participants (DiNardo et al., 2016).

Cognitive biases targeted by interventions

The most targeted cognitive biases were availability bias, search satisfying, and confirmation bias. All three biases are known to cause fixation on one leading diagnosis, which may lead to premature closure (Croskerry, 2003). Among the included articles, metacognitive principles delivered in the form of an educational intervention and/or checklist were shown to have been used to target all three biases in almost all instances. This is unsurprising, as metacognition encourages clinicians to reconsider the working hypothesis by analyzing the reasoning steps taken to arrive at that conclusion, which would help alleviate a fixation on one diagnosis. This finding, however, could also be the result of reporting bias.

Most interventions analyzed did not target specific cognitive biases or targeted a wide variety of biases. This might be a result of researchers attempting to develop solutions which can be applied to a wide array of situations, thus encouraging non-specific targeting of biases.

Ineffectiveness of some described interventions

5 studies reported at least one ineffective intervention, which collectively includes interventions from all three types of interventions discussed thus far. The authors' suggested reasons for the negative results vary greatly between studies and are largely speculative, but nevertheless may provide valuable insight into what future studies must avoid in designing their interventions.

Shimizu and colleagues reported one ineffective checklist intervention where their medical student participants unsuccessfully utilized a general debiasing checklist developed by Ely and colleagues (Ely et al., 2011; Shimizu et al., 2013). The authors suggested that due to the limited clinical experience that the participants had, heuristic thinking among the students may have been sparse. Therefore, a debiasing checklist would not have affected performance as cognitive bias occurs due to a fault in heuristics usage (Committee on Diagnostic Error in Health Care, 2015). Thus, if a debiasing intervention were to be tested on participants with relatively limited experience, such as medical students or junior residents, researchers should first ensure that participants demonstrate the use of heuristics or evidence of cognitive bias in the clinical situations they intend on examining.

The same debiasing checklist was also reported to be ineffective by Sibbald and colleagues (Sibbald et al., 2019). However, due to the use of simpler case vignettes, the baseline performance of both novices and experts was higher than in previous studies, which may explain the small effect size (Sibbald et al., 2013). Furthermore, a previously validated content-specific checklist also failed to demonstrate significant results in the same study (Sibbald et al., 2013, 2014), which could suggest a flaw in the study methodology which caused both checklists to be ineffective. Alternatively, since the studies examined diagnosis specifically in ECG, these results could show that debiasing checklists designed for general use are ineffective in specialized environments. High pre-intervention performance was also a suggested reason for the low effectiveness of another intervention implemented by Smith and Slack. The educational workshop demonstrated a low effect size, which was speculated to be due to the inconsistent assessment pre-workshop by attending physicians who were unfamiliar with the study evaluations.

Other instances of low intervention effectiveness include a debiasing checklist intervention reported by DiNardo and colleagues and a metacognitive education intervention by Payne. DiNardo noted highly variable use of the implemented tool and that it was used mostly by non-participant clinicians (i.e., attending physicians rather than resident participants). These factors may have contributed to the not significant findings (DiNardo et al., 2016). Payne reported that participants were distracted by graphical

representations of mental models included in the educational intervention and did not spend a great deal of time on the metacognitive feedback (Payne, 2012).

Concerns with current interventions and future research

While all three broad debiasing intervention types discussed demonstrated some promise in improving diagnostic accuracy, there are several problems which paint an unclear picture of their effectiveness in practice.

Most significantly, no intervention was evaluated in real clinical contexts. Thus, the useability and actual effectiveness of each intervention is unclear. Furthermore, all three debiasing interventions reported some instances of ineffectiveness; the exact reasonings for ineffectiveness are heterogeneous. Studies also examined the effectiveness of interventions immediately or soon after delivery of interventions. This experimental choice is logical for the purposes of research but puts the long-term effectiveness of interventions into question. Lastly, while it is immediately obvious why nearly all participants in debiasing interventions and initiatives are medical students and resident physicians, it is known that attending physicians are also vulnerable to cognitive bias (Croskerry, 2003). Therefore, the reported effectiveness of each intervention may not be accurate in attending physician populations. Moreover, a lack of attending physician participants in the included studies suggests a troubling lack of focus on the effects of cognitive biases in continued medical education.

Due to these flaws in the interventions examined in this review, we recommend that future studies should further examine the effectiveness of each intervention in 1) real clinical practice, 2) with participant populations which include attending physicians and other practicing clinicians responsible for diagnosis (i.e., NP and PA), 3) report the long term retention of educational materials or intervention usage, and 4) investigate reasoning behind ineffectiveness and low intervention usage. Future studies should also examine whether effectiveness and intervention usage in case vignettes translate directly to clinical practice.

Limitations

There are limitations to this review which should be acknowledged. First, articles were only included if the intervention specifically mentions that it is a debiasing intervention or reduces the negative impacts of cognitive biases in diagnostic decision-making. This criterion meant that some articles that may also reduce the negative impacts of cognitive biases were not included as they did not explicitly identify themselves as such. Second, due to the heterogeneity of the extracted data, we were unable to provide a formal meta-analysis. While we were able to categorize interventions into three different types, the details of each intervention and outcome measures between studies varied greatly. Third, while the inclusion criteria specified the inclusion of NP, PA, and attending physician participants, no NP and PA participants were included in the final analysis and a limited number of attending physician participants were included. Fourth, the scope of this review was limited by the number of databases chosen. A more comprehensive search could have led to more reliable and robust findings. Lastly, all studies analyzed decision-making in a vignette context which may not accurately reflect real-world practice.

Conclusion

In this review, we characterized and evaluated implemented debiasing interventions and educational initiatives on cognitive biases in diagnosis. Though the strategies discussed in this review reveal great potential in their effects on improving diagnostic accuracy, further research is required to develop more effective and practical solutions in clinical environments. We advocate for future research on the viability and useability of debiasing strategies in real clinical practice among practising physicians as well as students and resident physicians. In addition, researchers and medical educators should account for the time and cognitive restraints of real clinical practice and implement feedback from clinician participants to increase the useability of implemented debiasing strategies. Due to the prevalence and consequences of cognitive biases on diagnostic error, successful debiasing interventions have the potential to significantly reduce medical errors and improve patient outcomes (Graber et al., 2005; Okafor et al., 2016; Saber Tehrani et al., 2013).

Author Contributions

Both authors conceptualized the review, drafted the study protocol on PROSPERO, screened collected articles, extracted data, conducted quality assessment, and contributed equally to the creation of all figures and tables. AT formulated and executed the search strategy, conducted the formal analysis, drafted the abstract, introduction, discussion, and cover letter. MM drafted the methods and results, managed all citations, and revised all written materials. All data used and synthesized in this review are available from the corresponding author (AT) upon request.

Acknowledgements

The authors would like to thank Samantha Villamor for her assistance in the manuscript editing process.

References

- Armstrong, B. A., Dutescu, I. A., Nemoy, L., Bhavsar, E., Carter, D. N., Ng, K.-D., Boet, S., Trbovich, P., & Palter, V. (2022). Effect of the surgical safety checklist on provider and patient outcomes: a systematic review. *BMJ Quality & Safety*, *31*(6), 463–478. <https://doi.org/10.1136/bmjqs-2021-014361>
- Barnett, G. O., Cimino, J. J., Hupp, J. A., & Hoffer, E. P. (1987). DXplain. An evolving diagnostic decision-support system. *JAMA: The Journal of the American Medical Association*, *258*(1), 67–74. <https://doi.org/10.1001/jama.258.1.67>
- Bond, W. F., Deitrick, L. M., Arnold, D. C., Kostenbader, M., Barr, G. C., Kimmel, S. R., & WorriLOW, C. C. (2004). Using Simulation to Instruct Emergency Medicine Residents in Cognitive Forcing Strategies. *Academic Medicine*, *79*(5), 438–446. <https://doi.org/10.1097/00001888-200405000-00014>
- Chew, K. S., Durning, S. J., & van Merriënboer, J. J. (2016). Teaching metacognition in clinical decision-making using a novel mnemonic checklist: an exploratory study. *Singapore Medical Journal*, *57*(12), 694–700. <https://doi.org/10.11622/smedj.2016015>
- Committee on Diagnostic Error in Health Care. (2015). *Improving Diagnosis in Health Care* (E. P. Balogh, B. T. Miller, & J. R. Ball, Eds.). National Academies Press. <https://doi.org/10.17226/21794>
- Croskerry, P. (2003). The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them. *Academic Medicine*, *78*(8), 775–780. <https://doi.org/10.1097/00001888-200308000-00003>
- Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 2: impediments to and strategies for change. *BMJ Quality & Safety*, *22*(Suppl 2), 65–72. <https://doi.org/10.1136/bmjqs-2012-001713>
- DiNardo, D., Tilstra, S. A., Follansbee, W., Barnato, A. E., Zimmer, S. M., Farris, C., & McNeil, M. (2016). Combatting cognitive bias: A clinical reasoning curriculum for internal medicine residency education. *Journal of General Internal Medicine*, *31*(2), S161–S162.
- DiNardo, D., Tilstra, S. A., McNeil, M., Follansbee, W., Zimmer, S. M., Farris, C., Mamede, S., Schmidt, H., & Barnato, A. E. (2015). Evaluation of a diagnostic checklist for use in internal medicine resident education. *Journal of General Internal Medicine*, *30*, S152.
- Ely, J. W., Graber, M. L., & Croskerry, P. (2011). Checklists to Reduce Diagnostic Errors. *Academic Medicine*, *86*(3), 307–313. <https://doi.org/10.1097/ACM.0b013e31820824cd>
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition. *Perspectives on Psychological Science*, *8*(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Gawande, A. (2009). *The Checklist Manifesto - How to Get Things Right*. Metropolitan Books.
- Graber, M. (2003). Metacognitive Training to Reduce Diagnostic Errors. *Academic Medicine*, *78*(8), 781. <https://doi.org/10.1097/00001888-200308000-00004>
- Graber, M. L. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety*, *22*(Suppl 2), ii21–ii27. <https://doi.org/10.1136/bmjqs-2012-001615>
- Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic Error in Internal Medicine. *Archives of Internal Medicine*, *165*(13), 1493. <https://doi.org/10.1001/archinte.165.13.1493>

- Iyer, S., Goss, E., Browder, C., Paccione, G., & Arnsten, J. (2019). Development and evaluation of a clinical reasoning curriculum as part of an Internal Medicine Residency Program. *Diagnosis*, 6(2), 115–119. <https://doi.org/10.1515/dx-2018-0093>
- Jenkins, M. M., & Youngstrom, E. A. (2016). A randomized controlled trial of cognitive debiasing improves assessment and treatment selection for pediatric bipolar disorder. *Journal of Consulting and Clinical Psychology*, 84(4), 323–333. <https://doi.org/10.1037/ccp0000070>
- Jenkins, M., & Youngstrom, E. (2012). Cognitive de-biasing strategies significantly improve the assessment of pediatric bipolar disorder. *Neuropsychopharmacology*, 38(S1), S314–S446. <https://doi.org/10.1038/npp.2012.221>
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>
- Klein, G. (2008). Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 456–460. <https://doi.org/10.1518/001872008X288385>
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. Basic Books. <https://doi.org/10.5860/CHOICE.37-0239>
- Mamede, S., de Carvalho-Filho, M. A., de Faria, R. M. D., Franci, D., Nunes, M. do P. T., Ribeiro, L. M. C., Biegelmeyer, J., Zwaan, L., & Schmidt, H. G. (2020). Immunising’ physicians against availability bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Quality & Safety*, 29(7), 550–559. <https://doi.org/10.1136/bmjqs-2019-010079>
- Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M. J. P., van Saase, J. L. C. M., van Guldener, C., & Schmidt, H. G. (2010). Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents. *JAMA*, 304(11), 1198. <https://doi.org/10.1001/jama.2010.1276>
- McGuinness, L. A., & Higgins, J. P. T. (2021). Risk of bias VISualization (robvis): An R package and Shiny web app for visualizing risk of bias assessments. *Research Synthesis Methods*, 12(1), 55–61. <https://doi.org/10.1002/jrsm.1411>
- Myung, S. J., Kang, S. H., Phyoo, S. R., Shin, J. S., & Park, W. B. (2013). Effect of enhanced analytic reasoning on diagnostic accuracy: A randomized controlled study. *Medical Teacher*, 35(3), 248–250. <https://doi.org/10.3109/0142159X.2013.759643>
- Newman-Toker, D. E., Wang, Z., Zhu, Y., Nassery, N., Saber Tehrani, A. S., Schaffer, A. C., Yu-Moe, C. W., Clemens, G. D., Fanai, M., & Siegal, D. (2021). Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: toward a national incidence estimate using the “Big Three.” *Diagnosis*, 8(1), 67–84. <https://doi.org/10.1515/dx-2019-0104>
- Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The Causes of Errors in Clinical Reasoning. *Academic Medicine*, 92(1), 23–30. <https://doi.org/10.1097/ACM.0000000000001421>
- Ogdie, A. R., Reilly, J. B., Pang, W. G., Keddem, S., Barg, F. K., von Feldt, J. M., & Myers, J. S. (2012). Seen Through Their Eyes. *Academic Medicine*, 87(10), 1361–1367. <https://doi.org/10.1097/ACM.0b013e31826742c9>

- Okafor, N., Payne, V. L., Chathampally, Y., Miller, S., Doshi, P., & Singh, H. (2016). Using voluntary reports from physicians to learn from diagnostic errors in emergency medicine. *Emergency Medicine Journal*, 33(4), 245–252. <https://doi.org/10.1136/emered-2014-204604>
- O’Sullivan, E., & Schofield, S. (2018). Cognitive Bias in Clinical Medicine. *Journal of the Royal College of Physicians of Edinburgh*, 48(3), 225–232. <https://doi.org/10.4997/jrcpe.2018.306>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Payne, V. L. (2012). Effect of a metacognitive intervention on cognitive heuristic use during diagnostic reasoning. *Dissertation Abstracts International*, 72(11), 7060.
- Plackett, R., Kassianos, A. P., Kambouri, M., Kay, N., Mylan, S., Hopwood, J., Schartau, P., Gray, S., Timmis, J., Bennett, S., Valerio, C., Rodrigues, V., Player, E., Hamilton, W., Raine, R., Duffy, S., & Sheringham, J. (2020). Online patient simulation training to improve clinical reasoning: a feasibility randomised controlled trial. *BMC Medical Education*, 20(1), 245. <https://doi.org/10.1186/s12909-020-02168-4>
- Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist*, 47(9), 1102–1114. <https://doi.org/10.1037/0003-066X.47.9.1102>
- Reilly, J. B., Ogdie, A. R., von Feldt, J. M., & Myers, J. S. (2013). Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ Quality & Safety*, 22(12), 1044–1050. <https://doi.org/10.1136/bmjqs-2013-001987>
- Saber Tehrani, A. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E. (2013). 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank. *BMJ Quality & Safety*, 22(8), 672–680. <https://doi.org/10.1136/bmjqs-2012-001550>
- Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision Making*, 16(1), 138. <https://doi.org/10.1186/s12911-016-0377-1>
- Schiff, G. D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B. L., Elstein, A. S., Hasler, S., Kabongo, M. L., Krosnjar, N., Odwazny, R., Wisniewski, M. F., & McNutt, R. A. (2009). Diagnostic Error in Medicine. *Archives of Internal Medicine*, 169(20), 1881–1887. <https://doi.org/10.1001/archinternmed.2009.333>
- Shahid, F., Abbas, H., Mohmand, M. G., Gupta, N., Vehar, S., VenkataSubramani, M., McGervey, M., & Mehdi, A. (2017). Conscious observation and analysis of subconscious thinking (COAST): The harbinger of metacognitive approach to clinical reasoning. *Journal of General Internal Medicine*, 32(2), S664–S665.

- Sherbino, J., Dore, K. L., Siu, E., & Norman, G. R. (2011). The Effectiveness of Cognitive Forcing Strategies to Decrease Diagnostic Error: An Exploratory Study. *Teaching and Learning in Medicine*, 23(1), 78–84. <https://doi.org/10.1080/10401334.2011.536897>
- Sherbino, J., Kulasegaram, K., Howey, E., & Norman, G. (2014). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. *CJEM*, 16(01), 34–40. <https://doi.org/10.2310/8000.2013.130860>
- Shimizu, T., Matsumoto, K., & Tokuda, Y. (2013). Effects of the use of differential diagnosis checklist and general de-biasing checklist on diagnostic performance in comparison to intuitive diagnosis. *Medical Teacher*, 35(6), e1218–e1229. <https://doi.org/10.3109/0142159X.2012.742493>
- Sibbald, M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2013). Checklists improve experts' diagnostic decisions. *Medical Education*, 47(3), 301–308. <https://doi.org/10.1111/medu.12080>
- Sibbald, M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2014). Finding and fixing mistakes: do checklists work for clinicians with different levels of experience? *Advances in Health Sciences Education*, 19(1), 43–51. <https://doi.org/10.1007/s10459-013-9459-3>
- Sibbald, M., Sherbino, J., Ilgen, J. S., Zwaan, L., Blissett, S., Monteiro, S., & Norman, G. (2019). Debiasing versus knowledge retrieval checklists to reduce diagnostic error in ECG interpretation. *Advances in Health Sciences Education*, 24(3), 427–440. <https://doi.org/10.1007/s10459-019-09875-8>
- Singh, H. (2013). Diagnostic errors: moving beyond 'no respect' and getting ready for prime time. *BMJ Quality & Safety*, 22(10), 789–792. <https://doi.org/10.1136/bmjqs-2013-002387>
- Smith, B. W., & Slack, M. B. (2015). The effect of cognitive debiasing training among family medicine residents. *Diagnosis*, 2(2), 117–121. <https://doi.org/10.1515/dx-2015-0007>
- Yale, S. C., Cohen, S. S., Kliegman, R. M., & Bordini, B. J. (2022). A pause in pediatrics: implementation of a pediatric diagnostic time-out. *Diagnosis*, 9(3). <https://doi.org/10.1515/dx-2022-0010>

Figures and Tables

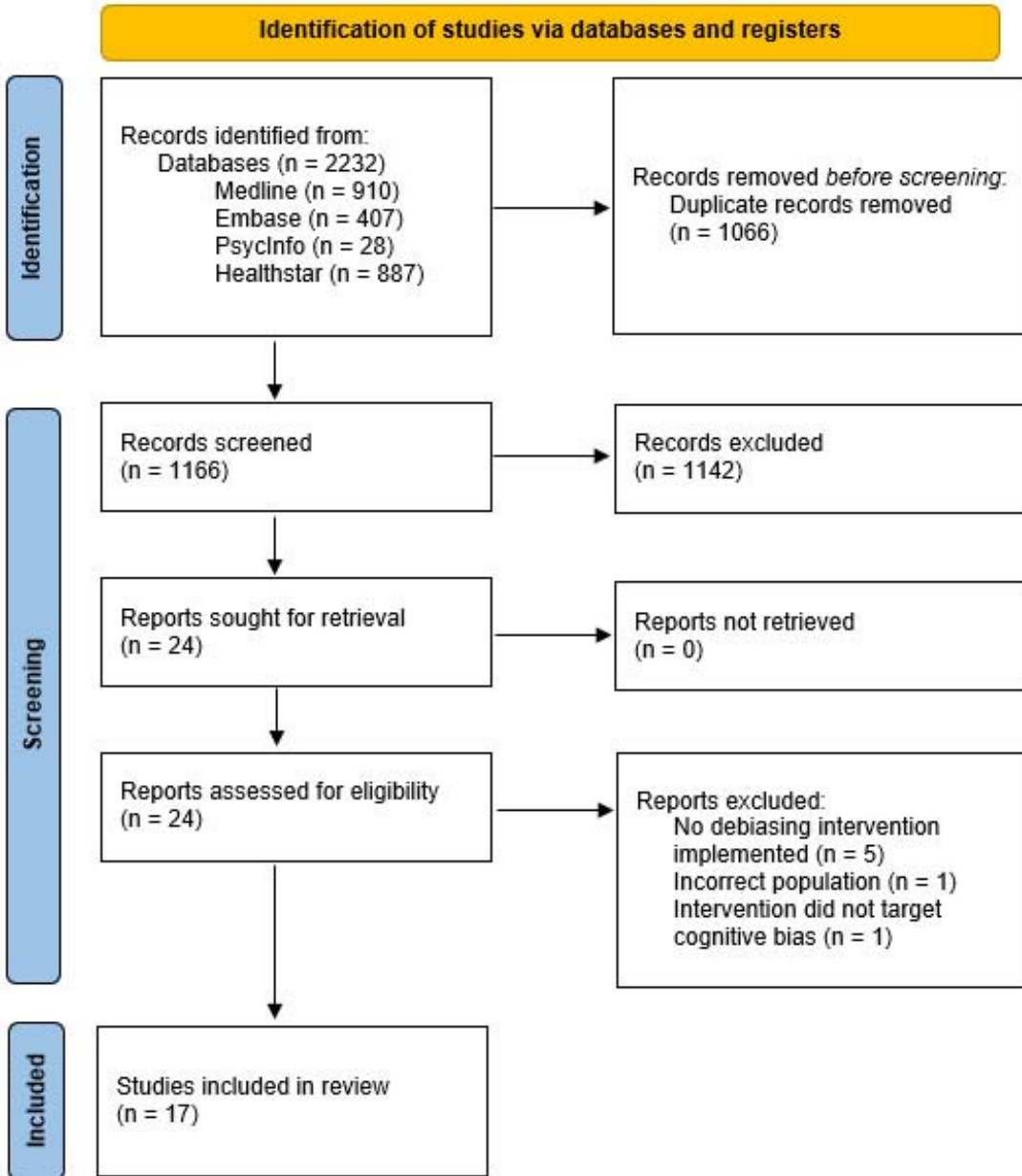


Figure 1. PRISMA flowchart of selected studies

Table 1. Details and effectiveness of debiasing interventions

Study	Intervention Type	Intervention Details	Effectiveness
Chew et al. (2016)	tool use	A mnemonic checklist was implemented to facilitate metacognitive reflection during vignette diagnosis and compared to control.	Use of checklist increase diagnostic performance in ECG vignettes.
DiNardo et al. (2015)	tool use	A reflective reasoning checklist was introduced after initial diagnosis of vignettes.	Use of checklist lead to increased diagnostic performance in vignettes.
DiNardo et al. (2016)	education of bias tool use	A workshop teaching concepts of clinical reasoning and cognitive biases was introduced alongside a diagnostic checklist.	Engagement with the workshop was not associated with change in diagnostic accuracy.
Iyer et al. (2019)	education of bias education of debiasing strategies	A clinical reasoning program educating residents on cognitive strategies, pitfalls, and biases of clinical reasoning and metacognitive principles was implemented.	Participant knowledge of clinical reasoning improved post-curriculum. Diagnostic performance was not evaluated.
Jenkins & Youngstrom (2012)	education of bias education of debiasing strategies	A workshop educating clinicians on strategies addressing common biases was delivered and compared to control.	Experimental participant showed significantly better diagnostic accuracy.
Jenkins & Youngstrom (2016)	education of bias education of debiasing strategies	A workshop educating clinicians on common biases and debiasing strategies (mnemonics, metacognition) was delivered and compared to control.	Experimental participant showed better overall judgement accuracy and fewer errors.
Mamede et al. (2010)	education of debiasing strategies	Participants were instructed on reflective reasoning techniques to reassess vignettes they had previously diagnosed with non-analytical thinking.	Reflection yielded more accurate diagnoses.
Mamede et al. (2020)	education of bias	Using example vignettes, diseases with similar presentations was compared/contrasted to prevent availability bias and compared to control.	Experimental participant were more accurate and immune to vignettes designed to induce bias.
Myung et al. (2013)	tool use	Intervention group was asked to list differential diagnoses and compare findings to each diagnosis before providing a diagnosis and compared to control.	Diagnostic accuracy significantly improved with analytical thinking.
Payne (2012) Experiment 2	education of debiasing strategies	Participants received feedback and educated in reasoning strategies and metacognitive thinking after the initial round of diagnoses.	Metacognition had no effect on diagnostic accuracy.
Plackett et al. (2018)	education of bias	Vignettes prompts participants to give justification behind each decision to engage analytical reasoning.	The intervention group felt more confident in their diagnoses and their information gathering skills improved. Diagnostic accuracy was not assessed.
Reilly et al. (2013)	education of bias	A curriculum educating junior residents on bias and diagnostic error was implemented to impact residents' ability to recognize bias during diagnosis.	Participants post-intervention scored higher in bias recognition/knowledge than both pre-intervention and r

			participant senior resident Diagnostic accuracy was assessed.
Shahid et al. (2017)	education of debiasing strategies	A “thinking aloud” strategy was implemented along with reflection and considering possible biases involved in diagnosis.	No outcome measure evaluated.
Sherbino et al. (2014)	education of debiasing strategies	Cognitive forcing strategies were taught to students (i.e. identifying multiple correct diagnoses for cases prone to search satisfying).	Intervention failed to reduce diagnostic errors.
Shimizu et al. (2013)	tool use	Two checklists (differential diagnosis and general debiasing) were implemented separately and compared to control.	The differential diagnosis checklist improved diagnostic accuracy in complex vignette cases while the general debiasing checklist was ineffective.
Sibbald et al. (2019)	tool use	Two checklists (specific ECG content and general debiasing) were implemented separately and compared to control.	In ECG vignettes, both checklists were ineffective in improving diagnostic accuracy.
Smith & Slack (2015)	education of bias	A workshop educating residents on recognizing bias and metacognition was delivered and pre/post vignette performance was recorded.	Diagnostic accuracy was unchanged post-workshop; participants' ability to formulate plans to detect bias was improved.

Table 2. Definitions of cognitive biases mentioned

Cognitive bias	Definition
Availability bias	The disposition to judge things as being more likely, or frequently occurring, if they readily come to mind.
Overconfidence	A belief that one knows more than they actually do or that they are more competent than they actually are.
Base rate neglect	The tendency to ignore the true prevalence of a disease, either inflating or reducing its base-rate, and distorting Bayesian reasoning.
Confirmation bias	The tendency to look for confirming evidence to support a diagnosis rather than look for disconfirming evidence to refute it
Anchoring bias	Locking on to salient features early in diagnosis and failing to adjust this first impression.
Search satisfying	Reflects the universal tendency to call off a search once something is found.
Diagnostic momentum	Once diagnostic labels are attached to patients they tend to become stickier and stickier.
Premature closure	Prematurely ending the clinical decision-making process and accepting a diagnosis before it can be verified.