**Abstract**

Objective

Overconfidence is an important source of medical error. Here we analyse experimental studies of confidence in medical diagnosis to identify factors affecting clinicians’ confidence in their diagnostic decisions, and how confidence impacts diagnosis and treatment.

Method

A scoping review of both medical and psychological literature was conducted. Articles were categorised according to methodology and clinical speciality. The findings were analysed thematically.

Data Sources

We systemically searched SCOPUS, MEDLINE, PsycINFO and Global Health then performed citation tracking within the references of these initial papers to identify missed articles.

Eligibility criteria

Studies were included if they reported quantitative results from an empirical study in which participants reported their confidence or certainty during a diagnostic decision. Studies comprised a broad set of medical subdisciplines.

Results

79 articles met the inclusion criteria. Across these articles, confidence was not found to be well-calibrated to true diagnostic accuracy regardless of clinician experience. We organised the articles under two main themes: the determinants of confidence (as opposed to objective diagnostic accuracy) and the uses of confidence later in the patient’s care pathway. On the former, confidence is found to be affected by several factors including patient case complexity, early diagnostic differentials, and context within the healthcare environment. Factors that affect confidence but not accuracy demonstrate how the two can be decoupled, resulting in overconfidence or underconfidence. On the latter theme, confidence is found to affect further patient testing, medication administration and referral rates among other factors.

Conclusions

Results from this review have implications for medical education and practice around diagnostic uncertainty and considerations of work from cognitive psychology. We propose a theoretical model of factors that affect diagnostic confidence/certainty and diagnostic accuracy. Such a model can inform future work on diagnosis, especially within medical education, on how appropriate confidence can be prompted and communicated amongst clinicians.

Word Count: 300

WHAT IS ALREADY KNOWN ON THIS TOPIC

There is extensive evidence of diagnostic error in most healthcare specialities and it has been suggested that cognitive biases, such as overconfidence, are causally linked with these errors. This study aims to synthesise past work on diagnostic confidence to understand the factors that contribute to clinicians’ confidence in their diagnoses, as well as how confidence informs other parts of patients’ care pathway.

WHAT THIS STUDY ADDS

As a result of this scoping review, we have identified that diagnostic confidence and accuracy have separate contributing factors related to the patient’s condition, the clinician making the diagnosis and the environmental context. We have also identified the pathways by which confidence, and misplaced confidence, can affect diagnosis and patient care. As a result of these findings, we introduce an integrative model of confidence throughout the patient care process.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

This study prompts future clinical work to focus not only on eliciting calibrated confidence but also on appropriate communication of uncertainty, which would have implications for medical education around how diagnostic reasoning is taught and for the workplace culture of healthcare environments. We also demonstrate the need to better model how confidence evolves over the course of a diagnostic decision rather than as a static quantity.

**INTRODUCTION**

Accurate medical diagnosis is crucial to high quality, safe patient care and is core part of a doctor’s job. Research on diagnosis has been grounded in the incidence of errors. A report from the US Institute of Medicine1 concluded that most patients will experience a diagnostic error within their lifetime. Around 32% of clinical errors have been found to be caused by clinician assessment, particularly the clinician’s failure to weigh up competing diagnoses2. Studies have also investigated the downstream consequences of diagnostic errors. Unnecessary treatments (or ‘overtreatment’) was estimated to cost the US healthcare system between 158 and 226 billion dollars in 20113. Diagnostic errors have also been found to lead to longer hospital stays and increased patient mortality4.

Heuristics are commonly used in diagnostic decisions that are necessarily often based on incomplete, imperfect information and made under time constraints. For example, making a diagnosis may involve considering a hypothesis as likely because the displayed symptoms correspond with a prototypical case of a particular condition5. A clinician may have recent experience of a patient with a particular condition and, when seeing another patient similar symptoms, will be more likely to diagnose that patient with the same condition 6. While these heuristics are often effective, associated cognitive biases are thought to be linked to diagnostic error. One cognitive bias that has been posited as a contributor to diagnostic error is overconfidence7, which may make an individual clinician fail to consider alternative diagnoses or lead a clinical team to be overly swayed by one individual’s opinion.

Confidence is defined in cognitive psychology as the subjective assessment of a decision’s quality or accuracy8. We refer to confidence as being ‘calibrated’ if it closely predicts objective accuracy (i.e., such that the individual is neither overconfident nor underconfident). In experimental studies, confidence sometimes exhibits impressive calibration with objective accuracy9, which is thought to reflect people’s ability to evaluate the quality of evidence on which they base their decisions10. But calibration is rarely perfect because confidence also depends on factors that do not directly correlate with accuracy, such as the time spent deliberating and the total amount of evidence considered (independent of the quality or consistency of this evidence)11,12, as well as the mood13, personality14, gender15 and status16 of the decision maker. The resulting under- and overconfidence matters: Overconfident decision makers leap to premature conclusions and ignore useful information or advice, while underconfident decision makers waste time collecting evidence that won’t improve their decisions17. Effective decision making in groups likewise depends on team members sharing calibrated information about their uncertainty: Confident team members tend to be listened to more, which can lead others astray if they are overconfident18. Conversely, underconfident team members may be ignored or may fail to share potentially useful information19.

These features of confidence highlight its potential importance in healthcare, as overconfidence can lead to insufficient consideration of diagnostic alternatives and inadequate care20. In the absence of objective feedback, confidence can be used as a marker of how likely someone is to be correct21. In medicine, a lack of clearly communicated feedback can cause clinicians to proceed as if they have received positive feedback. This means that they do not adequately update their internal model of the patient and hence they increase their confidence inappropriately, whether working individually or in teams22. Given growing interest in this topic, we conducted a scoping review to collate and synthesise the current work on studying diagnosis as a cognitive process, both in terms of what contributes to calibrated confidence and how judgements of confidence and certainty are utilised within the wider medical decision making process. The psychological literature treats these as different concepts, but they will treated as the same throughout this paper. Our full research questions can be found in Box 1.

**BOX 1: Scoping Review Research Questions (Preregistered)**

Primary questions:

* **RQ1:** How calibrated are the confidence/certainty judgements made during diagnostic decisions by clinicians relative to their actual accuracy?
* **RQ2:** How are confidence/certainty judgements utilised within the wider diagnostic decision process?

Subsidiary questions:

* **RQ3:** What are the prevalent ways in which diagnostic confidence and certainty are measured and operationalised as variables?
* **RQ4:** What strategies, tools or frameworks have been used to prompt better calibration of both confidence and certainty?
* **RQ5:** What types of empirical procedures/tasks are used to study confidence and certainty in diagnostic decisions? Do they come to different conclusions?
* **RQ6:** What are the discrepancies between the concepts/research questions studied in the context of confidence and certainty in the cognitive psychology literature and the medical diagnosis literature?
* **RQ7:** What areas of research are still underexplored within the context of medical diagnosis?

**METHODS**

**Search Strategy**

Our review protocol was preregistered on the Open Science Framework: <https://osf.io/wz5se>. We conducted a systematic scoping review of empirical studies of confidence and certainty in medical diagnosis using JBI’s PRIMSA-ScR Checklist for Scoping Reviews23. The search strategy was designed in cooperation with a subject specialist librarian at the University of Oxford’s Bodleian Libraries group. The search string was comprised of keywords to capture an intersection of four elements: confidence/certainty, medical diagnoses, decision making and a study population of medical staff/students (i.e., clinicians, physicians, doctors and medics). The full search terms can be found in Box 2. The databases SCOPUS, MEDLINE, PsycINFO and Global Health were searched during February 2024. Grey literature was searched using Finally, we hand-searched the citations of the included articles from these databases for further relevant articles via backward and forward scanning24,25.

**BOX 2: Search Terms**

(clinicians OR physicians OR doctors OR medics)

AND

( confiden\* OR uncertain\* OR certain\*)

AND

( diagnosis AND medical )

AND

( decision OR ( decision AND making ) OR decision-making )

**Study Selection**

The inclusion criteria for studies were as follows: (1) original empirical studies with quantitative results, (2) written in the English language, (3) experimental paradigm uses medical diagnostic decisions, (4) confidence or certainty is measured as a dependent variable. We exclude editorials, review papers and opinion papers, though we do include dissertations. We do not exclude any medical subdisciplines and do not exclude papers based on publication date. Identified articles were uploaded onto Rayyan (<https://rayyan.ai/>), which detects duplicate papers for manual checking and removal.

**Research Synthesis**

Papers selected for review were first categorised by their broad research methodology (e.g., patient vignettes, in situ questionnaires, etc.) and their medical population of study (e.g., medical students, general practitioners/primary care physicians/family doctors, medics etc.). We reviewed the experimental procedures to extract their key manipulations/independent variables. This included case complexity, use of a cognitive intervention and level of medical experience or knowledge of study participants. We also extracted dependent variables as they pertain to confidence/certainty and, where relevant, recording of both diagnostic differentials and information seeking. Each of the paper’s key findings were summarised and then all findings were categorised under recurring themes.

**RESULTS**

**Findings of Scoping Review**

The initial search returned a total of 3,332 articles. When applying the inclusion criteria described in the Study Selection section, 50 eligible articles were included. 439 articles were retrieved for review from the included articles’ citations. After applying both exclusions of duplicates and our inclusion criteria, 29 further articles were identified. This produced a total of 79 articles for inclusion and synthesis. The PRISMA diagram for this search process can be found in Figure 1.

A diagram of a flowchart

Description automatically generated

**FIGURE 1 – PRISMA Diagram of Literature Review.**

**Study Characteristics**

A summary of the included studies are found in Table 1. As shown in Figure 2, 36 of the 79 studies (46%) were published since 2019, indicating a recent surge of research interest in this field and the timeliness of a scoping review. The studies appeared in 59 different publications, including both medical and psychological journals, with journals related to medical education being most common (19 studies). Other research areas most represented were Primary Care/General Practice, Emergency Medicine and Nursing.

**Research Design**

Before discussing research themes that emerged from the included literature, we first set out the types of research designs. The included studies split roughly evenly between focusing on how confidence varies across individuals (XX studies) and on how confidence varies according to features of the patient case (XX studies). Of the former, most (34 studies, 43% of the sample) studied level of medical experience or training’s effect on confidence, either measured as a dependent variable or by recruiting participants in either a ‘novice’ or ‘experienced’ group. Of the latter, 19 studies (24%) manipulated the complexity or difficulty of the patient case. Finally, 10 studies (13%) investigated how diagnostic confidence varies the information presented or the opportunity to seek information.

Most of the studies (44 articles, 56% of the sample) used clinical patient text vignettes. For vignettes, there is an established ground truth in each case (unlike in situ studies involving real patients) to compare the participants’ responses to in order to gauge accuracy. As vignettes are quick and simple to administer, participants can complete several diagnoses during a single study such that both their confidence and accuracy can be averaged across cases. Other experimental methodologies include the use of imaging (e.g., ECG, X-Rays, MRI) for diagnosis, high-fidelity simulations (either using extended reality tools or using a patient mannequin), or questionnaires administered in situ to measure confidence during real patient cases as they are happening (Table 1). The preponderance of vignette studies is noteworthy given the finding from one study that nurses were both less accurate and less confident in a high-fidelity simulation compared to a paper-based vignette26, suggesting the need for caution when generalising less naturalistic paradigms (e.g. vignettes) to how clinicians would behave in their everyday medical practice.

Studies varied in how confidence and diagnostic accuracy were assessed. Studies mostly used a self-reported scale for confidence, (usually 1-10 or 1-100), as opposed to verbal expressions of confidence (e.g. “not sure”) or visual analogue scales. The use of such scales is common within cognitive psychology and such measured confidence values are often found to predict other behavioural variables of uncertainty, such as the tendency to seek further information or to opt out of making a decision27. Twenty-four studies (30%) allowed participants to input multiple diagnostic differentials rather than a single diagnosis. Confidence is then either measured for each differential or in the set of differentials as a whole.

In terms of accuracy, most studies prompt clinicians for a single diagnosis that is marked as correct or incorrect. Accuracy is easier to ‘mark’ for a single diagnosis, but it is less naturalistic to how clinicians may consider competing diagnoses in their everyday practice. Hence, 24 studies (30%) allowed participants to record multiple differentials in their diagnosis. An issue here is that accuracy then tends to be operationalised by looking at if a correct diagnosis is included in this set of multiple differentials. This means that clinicians are more likely to correct with more differentials and this method ignores how clinicians weigh up competing differentials. Hence, measures of how calibrated confidence judgements are to true diagnostic accuracy are heavily contingent on how diagnoses are recorded. This then has a bearing on how reliable findings on overconfidence/underconfidence are.

|  |  |  |  |
| --- | --- | --- | --- |
| **Publication Year** | | **Subdiscipline / Population** | |
| 1991-2000 | 10 | Primary Care / General Practice | 26 |
| 2001-2010 | 11 | Medical Students | 15 |
| 2011-2020 | 29 | Emergency Medicine | 10 |
| 2021- | 29 | Nursing | 6 |
|  |  | Pathology | 4 |
| **Study Environment/Context** | | Radiology | 4 |
| Textual Vignette | 44 | Other | 14 |
| Imaging Interpretation (e.g. ECG) | 20 |  |  |
| In Situ Questionnaires/Surveys | 13 | **Study Population Sample Size** | |
| High-Fidelity Simulation | 2 | < 100 | 44 |
|  |  | 101-200 | 20 |
| **Participant Experience Levels** | | 201-300 | 9 |
| Fixed Across Participants | 50 | > 300 | 6 |
| Multiple Experience Groups | 29 | **Total** | **79** |

**TABLE 1: Characteristics of Included Studies, including year of publication, study environment used and medical population (recruiting single or multiple levels of participant experience, medical subdiscipline, sample size).**

A graph of a number of years

Description automatically generated with medium confidence

**FIGURE 2 – Distribution of Papers by Publication Year**

**Emerging Research Themes**

Miscalibration of Confidence and Certainty Judgements with Objective Accuracy

Calibration is assessed by comparing confidence ratings with objective diagnostic accuracy: When clinicians rate 100% (or 50% or 60%, etc.) certainty in their diagnosis, are they in fact correct 100% (or 50%, 60%, etc.) of the time? Calibration is then an indirect measure that is calculated by comparing two other observed measures: confidence and accuracy. In our study sample, there was limited evidence of calibrated confidence judgements, with some studies reporting underconfidence26,28,29 and others overconfidence30-32. In order to examine these findings in more detail, we first consider factors that impact/promote confidence/uncertainty in diagnosis specifically (as this is directly measured in studies) before considering calibration itself.

Confidence is affected by a number of factors. Both confidence and case complexity seem to heavily interact with experience, with more experienced clinicians better able to pick up when a case is more complex/difficult and adjust their confidence accordingly33,34. Calibration was found to interact with experience such that underconfidence tended to be exhibited by less experienced clinicians (or students), whilst overconfidence was observed more for experienced clinicians. However, a difference in calibration across experience was not always observed in the results35,36. Looking at the link between calibration and experience alone may be too simplistic, and there are other aspects of experience that influence diagnoses. Experienced clinicians were found to be less likely to ‘distort’ neutral information to be in support of their reported diagnoses, indicating a lower tendency toward confirmation bias37.

A minor theme of past work has been a distinction between experience (operationalised as years of experience or role seniority) and knowledge (measured using standardised tests of medical knowledge). In medical students, the calibration of confidence judgements were found to improve with years of education but not with medical knowledge38. The aforementioned information ‘distortion’ was found to affect novice clinicians more37 and lower knowledge was found to be related to higher susceptibility to irrelevant, distracting features of a patient39. However, the latter study found that medical knowledge was not directly associated with confidence.

Studies have found that calibration is affected by the complexity or difficulty of the presented case40-42. When confidence judgements are not sensitive to the difficulty or complexity of the case, confidence stays fairly constant for difficult cases whilst accuracy decreases, leading to increased overconfidence (and decreased calibration). In past studies, complexity is manipulated by either presenting patient cases with more comorbid conditions41 or by showing more conflicting information about the patient to indicate multiple possible conditions33. Calibration can be improved by the presence of feedback during a training period43,44.

Contextual factors that pertain to the situated medical environment can also affect confidence, as found using more naturalistic paradigms. For example, clinicians may be constantly interrupted for other tasks45, especially during busier shifts where they have to manage more patients46 and may not be present for the sharing of information at handover 47. Studies that simulated these situations found lower diagnostic confidence. However they cannot assess the effect of contextual factors on calibration with diagnostic accuracy because they were conducted in-situ. though because the researchers for these studies do not obtain measures of accuracy due to working in situ, we can only determine how these contextual factors affect confidence rather than calibration.

**BOX 3: Papers on Imaging and Confidence**

There were also a subset of papers that find an increase in confidence when providing clinicians with specialised imaging for a patient when making diagnoses, be they MRI scans48-49, CT scans50, evacuation proctography51 or photos of wounds52. Meanwhile, another subset of papers use various forms of computer-aided decision support systems with the goal of improving confidence, with mixed results53-56. These results are perhaps not surprising, but do warrant addressing as per our inclusion criteria.

Interventions at the Point of Generating Differentials

The process of generating diagnostic differentials has been subject to experimental manipulations and interventions (such as early diagnostic suggestions) to investigate their effect on accuracy and confidence. One can imagine the applicability of this work, for instance when a clinician is transitioning care of a patient to another clinician and giving a handover of relevant information. A general theme of this work is that there is a tendency toward higher weighting of early information, indicating a primacy effect. Early diagnostic suggestions have been found to be highly influential in the subsequent decision process where clinicians find these suggestions difficult to ignore and have more confidence in them57-58. This also affects the breadth of differentials considered, with fewer differentials considered when provided with early suggestions59 and an underweighting of differentials if they were considered later in the diagnostic process60. Interventions aimed at mitigating this tendency by asking clinicians to explicitly consider alternatives, increased their accuracy and calibration61, or interventions prompting the consideration of the patient’s ‘red flags’ in diagnoses, increased confidence on simpler cases but not accuracy62. These interventions seemingly have to be explicit, as simply asking clinicians to reflect on their decision without guidance63,64 or participate in an educational training course65-66 does not seem to improve diagnostic accuracy and calibration.

Some studies also manipulated the manner in which information presented to clinicians during the diagnostic process affected confidence. Higher confidence was found when clinicians were presented with redundant patient information67 and when given all available patient information rather than having to gather information themselves68. Clinicians were also found to be more confident when presented with an Electronic Health Record of the patient alongside other information69 and when presented with the patient history first rather than out of order70, indicating a positive effect on confidence of a complete patient history available early on . An erroneous patient history has also been found to cue both novice and experienced clinicians to incorrect diagnoses whilst confidence remained relatively high, resulting in overall overconfidence71.

Uses of Confidence

With more naturalistic studies, it is possible to isolate ways in which confidence and certainty are utilised within the wider diagnostic process, especially as healthcare involves transitions of care between multiple clinicians and departments. Past work has attempted to establish a link between confidence and further seeking of patient information and tests with mixed results. US hospitalists with lower confidence were found to be linked to more test orders46 whilst more metacognitively aware pathologists (i.e. who tended to report confidence judgements that were closer to their true accuracy) were found to be more likely to request further tests when they were unsure36. Confidence has also been linked to prescribing medication, though overtreatment with unnecessary medications was found to be linked to both underconfidence72 and overconfidence32. Higher confidence has also been linked to referral rates to other specialists in other departments73 and to a lower willingness to admit mistakes29. One study found that whilst experienced clinicians were not more accurate in their diagnoses, they were more willing to change diagnoses and request more information74. Lower confidence has been found to result in less specific diagnoses for patients in situ75. Whilst past work on confidence has examined its role within groups (as discussed in the Introduction), only one included article looked at confidence in group decisions and found that a multidisciplinary panel was more confident and better calibrated than a single clinician76. We return to this later as a theme for future work.

Framework for Diagnostic Decisions

We synthesise the included findings into a theoretical framework to illustrate how various factors distinctly impact diagnostic confidence and accuracy. This framework is shown below in Figure 3. We especially note that the framework both summarises the existing research as described here and proposes directions for future research that has been relatively untapped by the included studies. Namely, the diagnostic decisions that are emulated experimentally view diagnosis as linear processes whilst future work could focus on better simulating the cyclical nature of diagnosis, such as how feedback is integrated into clinicians’ existing knowledge and how confidence and information seeking interact with each other. On the latter, past work has focused on how confidence is related to further testing and information seeking, but not how information seeking itself is related to confidence. Whilst past research has, as a whole, also hinted at a distinction between medical experience and knowledge as they pertain to confidence, there has been little work comparing and contrasting the two directly. This is especially pertinent given the social influence that seniority can have within a group, reducing the likelihood of more junior clinicians speaking up about potential errors in the presence of more experienced clinicians77.

A diagram of a medical procedure

Description automatically generated

**FIGURE 3: Overview of included research. This diagram describes the current state of included research. Contextual and environmental factors include access to MDTs, stress/busyness. The dark boxes show stages of the diagnostic decision process as they proceed over time. The black arrows indicate when a factor impacts the target. The green dashed arrows show links between factors that have been identified as recommendations for future work. EXPLAIN ALL BOXES/ARROWS, AND DIRECTION OF ARROW change to Evolution of a diagnostic decision. Patient early presentation/history. Can remove ongoing. Initial care, treatment instead of patient care. Feedback/learning/adaptation. Having different layers for individual, case, environment. Make clear how the decision is complex, evolutionary. Represent whether we know a lot/not that much about a link, where we have data or not.**

**DISCUSSION**

Our forward and backward citation search identified 37% of our sources, which is rather high but reveals a lot about the current state of the literature. We aimed to review studies on confidence during diagnoses and found a large number of studies through citation tracking. This review process reveals the broad applicability of confidence across medical subdisciplines. We also find that several studies measured confidence or certainty across a variety of different research methods (e.g. using ‘assessments’ or ‘interpretations’ as well as diagnostic decisions), but not as a primary variable for study. Future work would benefit from standardisation in terms of measuring confidence, as well as studying confidence itself in more detail. This is particularly important given the broad findings that confidence in diagnostic decisions tend to be miscalibrated. Given that confidence seems to interact with other factors like experience and the complexity of the patient case, future work could, in particular, prompt confidence in a manner that better reflects the individual and case-level factors. This review has also shown how confidence is associated with many aspects of the patient care process, illustrating just how it can influence clinicians’ behaviour and its importance to study, as well as its huge benefits to elucidate through further research.

**Strengths and Limitations**

Whereas previous reviews have focused on mapping instances of cognitive biases within medical errors78-80 or on medical uncertainty more broadly81,82, our review is the first to comprehensively map out the literature that links confidence and certainty to medical diagnoses in such a broad remit across medicine. Our work demonstrates that both errors and confidence are fruitful areas to study within diagnosis, as together they determine how calibrated a clinician is when expressing certainty/uncertainty. The breadth of the literature reviewed here, from a wide variety of medical subdisciplines, demonstrates how important confidence is as an area of study in terms of its broad applicability across healthcare. Given the wide range of subdisciplines that have studied confidence, there is then a need for a wider focus on confidence and metacognition within medical education. Findings from metacognition are already being used to inform educational practices outside medicine to improve students’ memory retention83,84.

**Implications and Future Clinical Research**

This scoping review shows the importance and the (particularly recent) surge in interest in diagnostic confidence. The recent interest in this field of work may be related to the increased focus on artificial intelligence in healthcare, particularly for diagnosis, and how important it is for these tools to convey uncertainty clearly. As such, there is interest in understanding where and how diagnostic uncertainty arises. Whilst confidence has been linked to diagnostic error in the past7, studying it requires insights from cognitive psychology to inform medical education and practice85. How clinicians evaluate their decisions contributes to their effectiveness, as an overconfident clinician may overlook diagnostic possibilities, delay treatment or ignore crucial information. Conversely, an underconfident clinician may be less likely to speak up in a group about potential errors77. Our review finds that confidence and accuracy are rarely aligned during diagnoses. Whilst cognitive interventions such as considering alternative diagnoses and guided reflections have been tested, there is yet to be a standardised cognitive framework to teach non-technical skills such as expression of uncertainty. Notably from these papers, miscalibration of confidence is not only a function of social and environmental factors, as it was also observed for vignette studies performed by individual participants. Such factors only serve to amplify systematic tendencies toward misaligned confidence/certainty.

Studying errors has tended to be the focus of past work, especially overconfidence as a source of errors. Our review provides evidence in support of this, as the included studies found evidence for overconfidence particularly when dealing with complex cases. Overconfidence was also found to be associated with overlooking differentials, ignoring important patient information and being less willing to admit mistakes. Hence, mitigating overconfidence is an important direction for future research. Underconfidence has received less attention, but is observed more in medical trainees26,28,29 and can lead to negative outcomes such as delayed treatment86. Interventions have be tested to improve confidence calibration (such as considering alternative diagnoses and guided reflection), but these have not been successful in this regard63,64. More work is then needed to design interventions to improve calibration.

Whilst we have reviewed a wide breadth of the available literature on confidence/certainty in diagnoses, there are still many avenues for future work. Namely, the majority of the studies presented here do not study diagnosis as a constantly ongoing and evolving process. This is likely a function of the methodologies used where participants are asked for diagnoses at a particular point in time, usually after reading all the available information on a patient. In reality, diagnoses feature a back and forth between seeking information and evaluating that information in the context of currently considered diagnostic possibilities. Without this consideration in experimental design, researchers cannot study the diagnostic process up until the point of expressing confidence. Past work has viewed confidence as it effects the subsequent diagnostic process (e.g. medications, testing) after this confidence is expressed, rather than what causes this sense of confidence in the first place. One way to study this is using more naturalistic, in situ methodologies that are more analogous to everyday medical practice. However, interrupting clinicians in real time to report their diagnostic thinking can be a distraction and potentially a safety risk. Future research should first focus on utilising other methodologies for capturing the diagnostic reasoning of clinicians as it evolves with time and the receipt of new information. This could include getting clinicians to think aloud as they make diagnoses87 or using a visual representation of clinicians’ thought processes to capture paths and sources of diagnoses61. The use of high-fidelity simulations is also useful for emulating the pressure and work environment of the clinician (which may affect decision making), as well as providing an actual ‘patient’ to observe (unlike in textual vignettes). Use of such paradigms would also improve the generalisability of results.

Most of the reviewed studies investigate confidence in individual clinicians. However, diagnosis and treatment decisions are often made by teams rather than individuals, particularly in secondary care settings. Evidence from organisational psychology indicates that group decisions depend critically on communicated confidence and uncertainty19: Overconfident team members can anchor a group on an incorrect decision88. Conversely, underconfident team members may fail to share important information that is unknown to the rest of the group, exacerbating the problems of ‘hidden information’ and ‘shared information bias’. In addition, future work could seek to identify how the certainty held by individuals differs from the certainty that one has to communicate to others (be they patients or other clinicians) in their diagnoses. Clinicians may modify how they communicate certainty with others, especially given the collaborative nature of healthcare and social benefits of communicating opinions with confidence in order to be listened to in a group29. The papers in this review have also not looked at individual differences in expressions of confidence, where past work from cognitive psychology has found individual systematic tendencies toward higher or lower confidence89,90. Hence, whilst task-level or environmental factors affect confidence and calibration, individual clinicians may also have trait-level factors that are predictive too.

**CONCLUSIONS**

This scoping review indicates that confidence is frequently not calibrated to accuracy during diagnostic decisions. We also found across the literature that different factors affect confidence and accuracy separately, which may help to explain why such instances of overconfidence or underconfidence are observed. Finally we identified several papers that underscore how confidence affects the subsequent care pathway of patients. Taken together, these findings have implications for how diagnostic certainty and confidence should be studied in future clinical work, including the role that information gathering and interpretation has on diagnoses and usage of naturalistic paradigms.

**REFERENCES**

1. McGlynn EA, McDonald KM, Cassel CK. Measurement is essential for improving diagnosis and reducing diagnostic error: a report from the Institute of Medicine. [Jama](https://pubmed.ncbi.nlm.nih.gov/26571126/). 2015 Dec 15;314(23):2501-2.
2. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, Elstein AS, Hasler S, Kabongo ML, Krosnjar N, Odwazny R. Diagnostic error in medicine: analysis of 583 physician-reported errors. [Archives of internal medicine](https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/1108559). 2009 Nov 9;169(20):1881-7.
3. Berwick DM, Hackbarth AD. Eliminating waste in US health care. [Jama](https://jamanetwork.com/journals/jama/article-abstract/1148376). 2012 Apr 11;307(14):1513-6.
4. Hautz WE, Kämmer JE, Hautz SC, Sauter TC, Zwaan L, Exadaktylos AK, Birrenbach T, Maier V, Müller M, Schauber SK. Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room. [Scandinavian journal of trauma, resuscitation and emergency medicine](https://link.springer.com/article/10.1186/s13049-019-0629-z). 2019 Dec;27:1-2.
5. Restrepo D, Armstrong KA, Metlay JP. Annals clinical decision making: avoiding cognitive errors in clinical decision making. [Annals of internal medicine](https://www.acpjournals.org/doi/abs/10.7326/M19-3692). 2020 Jun 2;172(11):747-51.
6. Mamede S, van Gog T, van den Berge K, Rikers RM, van Saase JL, van Guldener C, Schmidt HG. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. [Jama](https://jamanetwork.com/journals/jama/article-abstract/186585). 2010 Sep 15;304(11):1198-203.
7. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. [The American journal of medicine](https://www.sciencedirect.com/science/article/pii/S0002934308000405). 2008 May 1;121(5):S2-3.
8. Fleming SM, Daw ND. Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. [Psychological review](https://psycnet.apa.org/fulltext/2016-60724-003.html). 2017 Jan;124(1):91.
9. Boldt A, Yeung N. Shared neural markers of decision confidence and error detection. Journal of Neuroscience. 2015 Feb 25;35(8):3478-84.
10. Shekhar M, Rahnev D. How do humans give confidence? A comprehensive comparison of process models of perceptual metacognition. Journal of Experimental Psychology: General. 2023 Dec 14.
11. Kiani R, Corthell L, Shadlen MN. Choice certainty is informed by both evidence and decision time. Neuron. 2014 Dec 17;84(6):1329-42.
12. Ko YH, Feuerriegel D, Turner W, Overhoff H, Niessen E, Stahl J, Hester R, Fink GR, Weiss PH, Bode S. Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements. Cognition. 2022 Aug 1;225:105125.
13. Rouault M, Seow T, Gillan CM, Fleming SM. Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biological psychiatry. 2018 Sep 15;84(6):443-51.
14. Schaefer PS, Williams CC, Goodie AS, Campbell WK. Overconfidence and the big five. Journal of research in Personality. 2004 Oct 1;38(5):473-80.
15. Syzmanowicz A, Furnham A. Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses. Learning and individual Differences. 2011 Oct 1;21(5):493-504.
16. See KE, Morrison EW, Rothman NB, Soll JB. The detrimental effects of power on confidence, advice taking, and accuracy. Organizational behavior and human decision processes. 2011 Nov 1;116(2):272-85.
17. Desender K, Boldt A, Yeung N. Subjective confidence predicts information seeking in decision making. Psychological science. 2018 May;29(5):761-78.
18. Zarnoth P, Sniezek JA. The social influence of confidence in group decision making. Journal of Experimental Social Psychology. 1997 Jul 1;33(4):345-66.
19. Silver I, Mellers BA, Tetlock PE. Wise teamwork: Collective confidence calibration predicts the effectiveness of group discussion. Journal of Experimental Social Psychology. 2021 Sep 1;96:104157.
20. Kovacs RJ, Lagarde M, Cairns J. Overconfident health workers provide lower quality healthcare. Journal of Economic Psychology. 2020 Jan 1;76:102213.
21. Price PC, Stone ER. Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. [Journal of Behavioral Decision Making](https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.460). 2004 Jan;17(1):39-57.
22. Jaspan O, Wysocka A, Sanchez C, Schweitzer AD. Improving the relationship between confidence and competence: implications for diagnostic radiology training from the psychology and medical literature. [Academic Radiology](https://www.sciencedirect.com/science/article/pii/S1076633220306991). 2022 Mar 1;29(3):428-38.
23. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil, H. Chapter 11: Scoping Reviews (2020 version). In: Aromataris E, Munn Z (Editors). , JBI, 2020. Available JBI Manual for Evidence Synthesis from <https://synthesismanual.jbi.glo>
24. Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence‐informed management knowledge by means of systematic review. British journal of management. 2003 Sep;14(3):207-22.
25. Webster J, Watson RT. Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly. 2002 Jun 1:xiii-xiii.
26. Yang H, Thompson C, Bland M. The effect of clinical experience, judgment task difficulty and time pressure on nurses’ confidence calibration in a high fidelity clinical simulation. BMC medical informatics and decision making. 2012 Dec;12:1-9.
27. Gherman S, Philiastides MG. Neural representations of confidence emerge from the process of decision formation during perceptual choices. Neuroimage. 2015 Feb 1;106:134-43.
28. Mann D. The Relationship between Diagnostic Accuracy and Confidence in Medical Students. [ERIC](https://eric.ed.gov/?id=ED358110)
29. Brezis M, Orkin-Bedolach Y, Fink D, Kiderman A. Does Physician's Training Induce Overconfidence That Hampers Disclosing Errors?. [Journal of Patient Safety](https://journals.lww.com/journalpatientsafety/abstract/2019/12000/does_physician_s_training_induce_overconfidence.10.aspx). 2019 Dec 1;15(4):296-8.
30. Friedman C, Gatti G, Elstein A, Franz T, Murphy G, Wolf F. Are clinicians correct when they believe they are correct? Implications for medical decision support. In MEDINFO 2001 (pp. 454-458). IOS Press.
31. Fernández‐Aguilar C, Martín‐Martín JJ, Minué Lorenzo S, Fernández Ajuria A. Use of heuristics during the clinical decision process from family care physicians in real conditions. Journal of Evaluation in Clinical Practice. 2022 Feb;28(1):135-41.
32. Garbayo LS, Harris DM, Fiore SM, Robinson M, Kibble JD. A metacognitive confidence calibration (MCC) tool to help medical students scaffold diagnostic reasoning in decision-making during high-fidelity patient simulations. Advances in Physiology Education. 2023 Mar 1;47(1):71-81.
33. Brannon LA, Carson KL. Nursing expertise and information structure influence medical decision making. [Applied Nursing Research](https://www.sciencedirect.com/science/article/pii/S0897189703000788). 2003 Nov 1;16(4):287-90.
34. Tabak N, Bar-Tal Y, Cohen-Mansfield J. Clinical decision making of experienced and novice nurses. [Western Journal of Nursing Research](https://journals.sagepub.com/doi/abs/10.1177/019394599601800505). 1996 Oct;18(5):534-47.
35. Yang H, Thompson C. Nurses’ risk assessment judgements: A confidence calibration study. Journal of Advanced Nursing. 2010 Dec;66(12):2751-60.
36. Clayton DA, Eguchi MM, Kerr KF, Miyoshi K, Brunyé TT, Drew T, Weaver DL, Elmore JG. Are Pathologists Self-Aware of Their Diagnostic Accuracy? Metacognition and the Diagnostic Process in Pathology. Medical Decision Making. 2023 Feb;43(2):164-74.
37. Kostopoulou O, Russo JE, Keenan G, Delaney BC, Douiri A. Information distortion in physicians’ diagnostic judgments. Medical Decision Making. 2012 Nov;32(6):831-9.
38. Hautz WE, Schubert S, Schauber SK, Kunina‐Habenicht O, Hautz SC, Kämmer JE, Eva KW. Accuracy of self‐monitoring: does experience, ability or case difficulty matter?. [Medical education](https://asmepublications.onlinelibrary.wiley.com/doi/abs/10.1111/medu.13801). 2019 Jul;53(7):735-44.
39. Mamede S, Zandbergen A, de Carvalho-Filho MA, Choi G, Goeijenbier M, van Ginkel J, Zwaan L, Paas F, Schmidt HG. Role of knowledge and reasoning processes as predictors of resident physicians’ susceptibility to anchoring bias in diagnostic reasoning: a randomised controlled experiment. BMJ Quality & Safety. 2024 Feb 16.
40. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians’ diagnostic accuracy, confidence, and resource requests: a vignette study. JAMA internal medicine. 2013 Nov 25;173(21):1952-8.
41. Hausmann D, Kiesel V, Zimmerli L, Schlatter N, von Gunten A, Wattinger N, Rosemann T. Sensitivity for multimorbidity: The role of diagnostic uncertainty of physicians when evaluating multimorbid video case-based vignettes. PloS one. 2019 Apr 10;14(4):e0215049.
42. Li S, Zheng J, Lajoie SP. The relationship between cognitive engagement and students’ performance in a simulation-based training environment: an information-processing perspective. Interactive Learning Environments. 2023 Apr 3;31(3):1532-45.
43. Kuhn J, van den Berg P, Mamede S, Zwaan L, Bindels P, van Gog T. Improving medical residents’ self-assessment of their diagnostic accuracy: does feedback help?. [Advances in Health Sciences Education](https://link.springer.com/article/10.1007/s10459-021-10080-9). 2022 Mar;27(1):189-200.
44. Staal J, Katarya K, Speelman M, Brand R, Alsma J, Sloane J, Van den Broek WW, Zwaan L. Impact of performance and information feedback on medical interns' confidence–accuracy calibration. [Advances in Health Sciences Education](https://link.springer.com/article/10.1007/s10459-023-10252-9). 2023 Jun 17:1-7.
45. Soares III WE, Price LL, Prast B, Tarbox E, Mader TJ, Blanchard R. Accuracy screening for ST elevation myocardial infarction in a task-switching simulation. Western Journal of Emergency Medicine. 2019 Jan;20(1):177.
46. Gupta AB, Greene MT, Fowler KE, Chopra VI. Associations Between Hospitalist Shift Busyness, Diagnostic Confidence, and Resource Utilization: A Pilot Study. Journal of Patient Safety. 2023 Oct 1;19(7):447-52.
47. Bergl PA, Shukla N, Shah J, Khan M, Patel JJ, Nanchal RS. Factors influencing diagnostic accuracy among intensive care unit clinicians–an observational study. Diagnosis. 2024 Feb 19;11(1):31-9.
48. Mackenzie R, Dixon AK, Keene GS, Hollingworth W, Lomas DJ, Villar RN. Magnetic resonance imaging of the knee: assessment of effectiveness. Clinical radiology. 1996 Apr 1;51(4):245-50.
49. Albrechtsen SS, Riis RG, Amiri M, Tanum G, Bergdal O, Blaabjerg M, Simonsen CZ, Kondziella D. Impact of MRI on decision-making in ICU patients with disorders of consciousness. Behavioural Brain Research. 2022 Mar 12;421:113729.
50. Abujudeh HH, Kaewlai R, McMahon PM, Binder W, Novelline RA, Gazelle GS, Thrall JH. Abdominopelvic CT increases diagnostic certainty and guides management decisions: a prospective investigation of 584 patients in a large academic medical center. American Journal of Roentgenology. 2011 Feb;196(2):238-43.
51. Harvey CJ, Halligan S, Bartram CI, Hollings N, Sahdev A, Kingston K. Evacuation proctography: a prospective study of diagnostic and therapeutic effects. Radiology. 1999 Apr;211(1):223-7.
52. Sanger PC, Simianu VV, Gaskill CE, Armstrong CA, Hartzler AL, Lordon RJ, Lober WB, Evans HL. Diagnosing surgical site infection using wound photography: a scenario-based study. Journal of the American College of Surgeons. 2017 Jan 1;224(1):8-15.
53. Hillson SD, Connelly DP, Liu Y. The effects of computer-assisted electrocardiographic interpretation on physicians' diagnostic decisions. Medical Decision Making. 1995 Jun;15(2):107-12.
54. Berner ES, Maisiak RS. Influence of case and physician characteristics on perceptions of decision support systems. Journal of the American Medical Informatics Association. 1999 Sep 1;6(5):428-34.
55. Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artificial intelligence in medicine. 2005 Jan 1;33(1):25-30.
56. Neugebauer M, Ebert M, Vogelmann R. A clinical decision support system improves antibiotic therapy for upper urinary tract infection in a randomized single-blinded study. BMC Health Services Research. 2020 Dec;20:1-0.
57. Kämmer JE, Schauber SK, Hautz SC, Stroben F, Hautz WE. Differential diagnosis checklists reduce diagnostic error differentially: a randomised experiment. Medical education. 2021 Oct;55(10):1172-82.
58. Kourtidis P, Nurek M, Delaney B, Kostopoulou O. Influences of early diagnostic suggestions on clinical reasoning. Cognitive Research: Principles and Implications. 2022 Dec 15;7(1):103.
59. Staal J, Speelman M, Brand R, Alsma J, Zwaan L. Does a suggested diagnosis in a general practitioners’ referral question impact diagnostic reasoning: an experimental study. BMC Medical Education. 2022 Apr 8;22(1):256.
60. Eva WK. The influence of differentially processing evidence on diagnostic decision-making (Doctoral dissertation).
61. Feyzi-Behnagh R, Azevedo R, Legowski E, Reitmeyer K, Tseytlin E, Crowley RS. Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system. Instructional science. 2014 Mar;42:159-81.
62. Chartan C, Singh H, Krishnamurthy P, Sur M, Meyer A, Lutfi R, Stark J, Thammasitboon S. Isolating red flags to enhance diagnosis (I-RED): an experimental vignette study. International Journal for Quality in Health Care. 2019 Oct 31;31(8):G97-102.
63. Lambe KA, Hevey D, Kelly BD. Guided reflection interventions show no effect on diagnostic accuracy in medical students. Frontiers in psychology. 2018 Nov 23;9:285916.
64. Costa Filho GB, Moura AS, Brandão PR, Schmidt HG, Mamede S. Effects of deliberate reflection on diagnostic accuracy, confidence and diagnostic calibration in dermatology. Perspectives on Medical Education. 2019 Aug 1;8:230-6.
65. Benvenuto-Andrade C, Dusza SW, Hay JL, Agero AL, Halpern AC, Kopf AW, Marghoob AA. Level of confidence in diagnosis: clinical examination versus dermoscopy examination. Dermatologic surgery. 2006 May 1;32(5):738-44.
66. Kuhn J, Mamede S, van den Berg P, Zwaan L, van Peet P, Bindels P, van Gog T. Learning deliberate reflection in medical diagnosis: does learning-by-teaching help?. Advances in Health Sciences Education. 2023 Mar;28(1):13-26.
67. Heller RF, Saltzstein HD, Caspe WB. Heuristics in medical and non-medical decision-making. [The Quarterly Journal of Experimental Psychology Section A](https://journals.sagepub.com/doi/abs/10.1080/02724989243000019). 1992 Feb;44(2):211-35.
68. Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. [Medical Decision Making](https://journals.sagepub.com/doi/abs/10.1177/0272989x9101100401). 1991 Dec;11(4):233-9.
69. Ben-Assuli O, Sagi D, Leshno M, Ironi A, Ziv A. Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study. [Journal of biomedical informatics](https://www.sciencedirect.com/science/article/pii/S1532046415000477). 2015 Jun 1;55:31-40.
70. Tio RA, Carvalho Filho MA, de Menezes Mota MF, Santanche A, Mamede S. The Effect of Information Presentation Order on Residents' Diagnostic Accuracy of Online Simulated Patients With Chest Pain. [Journal of graduate medical education](https://meridian.allenpress.com/jgme/article/14/4/475/484936/The-Effect-of-Information-Presentation-Order-on). 2022 Aug 1;14(4):475-81.
71. Fawver B, Thomas JL, Drew T, Mills MK, Auffermann WF, Lohse KR, Williams AM. Seeing isn’t necessarily believing: Misleading contextual information influences perceptual-cognitive bias in radiologists. [Journal of Experimental Psychology: Applied](https://psycnet.apa.org/record/2020-28446-001). 2020 Dec;26(4):579.
72. Levin PD, Idrees S, Sprung CL, Weissman C, Weiss Y, Moses AE, Benenson S. Antimicrobial use in the ICU: indications and accuracy—an observational trial. [Journal of hospital medicine](https://shmpublications.onlinelibrary.wiley.com/doi/abs/10.1002/jhm.1964). 2012 Nov;7(9):672-8.
73. Calman NS, Hyman RB, Licht W. Variability in consultation rates and practitioner level of diagnostic certainty. [J Fam Pract](https://www.researchgate.net/profile/Neil-Calman/publication/21858297_Variability_in_Consultation_Rates_and_Practitioner_Level_of_Diagnostic_Certainty/links/57961ea008aec89db7b84cfd/Variability-in-Consultation-Rates-and-Practitioner-Level-of-Diagnostic-Certainty.pdf). 1992 Jul 1;35(1):31-8.
74. Krupat E, Wormwood J, Schwartzstein RM, Richards JB. Avoiding premature closure and reaching diagnostic accuracy: some key predictive factors. Medical education. 2017 Nov;51(11):1127-37.
75. Hageman MG, Bossen JK, King JD, Ring D. Surgeon confidence in an outpatient setting. [Hand](https://journals.sagepub.com/doi/abs/10.1007/s11552-013-9533-6). 2013 Dec;8(4):430-3.
76. Thorlacius-Ussing G, Bruun M, Gjerum L, Frederiksen KS, Rhodius-Meester HF, Van Der Flier WM, Waldemar G, Hasselbalch SG. Comparing a single clinician versus a multidisciplinary consensus conference approach for dementia diagnostics. Journal of Alzheimer's Disease. 2021 Jan 1;83(2):741-51.
77. Hémon B, Michinov E, Guy D, Mancheron P, Scipion A. Speaking up about errors in routine clinical practice: a simulation-based intervention with nursing students. Clinical Simulation in Nursing. 2020 Aug 1;45:32-41.
78. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care—a systematic review. Family practice. 2008 Dec 1;25(6):400-13.
79. Graber ML, Kissam S, Payne VL, Meyer AN, Sorensen A, Lenfestey N, Tant E, Henriksen K, LaBresh K, Singh H. Cognitive interventions to reduce diagnostic error: a narrative review. BMJ quality & safety. 2012 Jul 1;21(7):535-57.
80. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC medical informatics and decision making. 2016 Dec;16:1-4.
81. Hall KH. Reviewing intuitive decision‐making and uncertainty: the implications for medical education. Medical education. 2002 Mar;36(3):216-24.
82. Bhise V, Rajan SS, Sittig DF, Morgan RO, Chaudhary P, Singh H. Defining and measuring diagnostic uncertainty in medicine: a systematic review. Journal of general internal medicine. 2018 Jan;33:103-15.
83. Dunlosky J. Strengthening the student toolbox: Study strategies to boost learning. American Educator. 2013;37(3):12-21.
84. Putnam AL, Sungkhasettee VW, Roediger III HL. Optimizing learning in college: Tips from cognitive psychology. Perspectives on Psychological Science. 2016 Sep;11(5):652-60.
85. Wears RL. Diagnosing diagnosis. Annals of Emergency Medicine. 2014 Dec 1;64(6):586-7.
86. Thompson C, Cullum N, McCaughan D, Sheldon T, Raynor P. Nurses, information use, and clinical decision making—the real world potential for evidence-based decisions in nursing. Evidence-based nursing. 2004 Jul 1;7(3):68-72.
87. Coderre S, Mandin HH, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. Medical education. 2003 Aug;37(8):695-703.
88. Mahmoodi A, Bang D, Olsen K, Zhao YA, Shi Z, Broberg K, Safavi S, Han S, Nili Ahmadabadi M, Frith CD, Roepstorff A. Equality bias impairs collective decision-making across cultures. Proceedings of the National Academy of Sciences. 2015 Mar 24;112(12):3835-40.
89. Ais J, Zylberberg A, Barttfeld P, Sigman M. Individual consistency in the accuracy and distribution of confidence judgments. Cognition. 2016 Jan 1;146:377-86.
90. Navajas J, Hindocha C, Foda H, Keramati M, Latham PE, Bahrami B. The idiosyncratic nature of confidence. Nature human behaviour. 2017 Nov;1(11):810-8.

**TABLE 2: Full Table of Included Studies (Supplemental)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Author(s) | Title | Year | Discipline | Methodology | Measure of Confidence |
| Neugebauer, M.; Ebert, M.; Vogelmann, R. | A clinical decision support system improves antibiotic therapy for upper urinary tract infection in a randomized single-blinded study. | 2020 | Medical Doctors (Internal Medicine) | Fictive Paper Case | Confidence in Diagnosis (%) |
| Sanger, P. C.; Simianu, V. V.; Gaskill, C. E.; Armstrong, C. A. L.; Hartzler, A. L.; Lordon, R. J.; Lober, W. B.; Evans, H. L. | Diagnosing surgical site infection using wound photography: a scenario-based study. | 2017 | Members of Surgical Infection Society | 5 online scenarios | Confidence in diagnosis (1-10) |
| Levin, P. D.; Idrees, S.; Sprung, C. L.; Weissman, C.; Weiss, Y.; Moses, A. E.; Benenson, S. | Antimicrobial use in the ICU: Indications and accuracy - an observational trial. | 2012 | ICU | Observational in ICU | Certainty of presence of infection when starting patients on antimicrobials |
| Fernandez-Aguilar, Carmen; Martin-Martin, Jose Jesus; Minue Lorenzo, Sergio; Fernandez Ajuria, Alberto | Use of heuristics during the clinical decision process from family care physicians in real conditions. | 2022 | Primary Care | Real patients presenting with dyspnoea | 0-100% scale confidence in diagnosis |
| Heller, Rachael F; Saltzstein, Herbert D; Caspe, William B | Heuristics in medical and non-medical decision-making. | 1992 | Paediatric residents | Medical and non-medical problems | 0-100% scale confidence in diagnosis |
| Mackenzie, R; Dixon, A K; Keene, G S; Hollingworth, W; Lomas, D J; Villar, R N | Magnetic resonance imaging of the knee: assessment of effectiveness. | 1996 | Radiology | Observation of knee MRI patients | 5 point visual analogue confidence scale |
| Mamede, S.; Zandbergen, A.; De Carvalho-Filho, M.A.; Choi, G.; Goeijenbier, M.; Van Ginkel, J.; Zwaan, L.; Paas, F.; Schmidt, H.G. | Role of knowledge and reasoning processes as predictors of resident physicians' susceptibility to anchoring bias in diagnostic reasoning: A randomised controlled experiment | 2024 | Internal Medicine | 6 clinical vignettes (with vs without salient distracting features) | Confidence in diagnosis |
| Staal, J.; Speelman, M.; Brand, R.; Alsma, J.; Zwaan, L. | Does a suggested diagnosis in a general practitioners’ referral question impact diagnostic reasoning: an experimental study | 2022 | Internal Medicine | 6 cases formatted as GP referral letters | 0-10 confidence in diagnosis |
| Küper, A.; Lodde, G.; Livingstone, E.; Schadendorf, D.; Krämer, N. | Mitigating cognitive bias with clinical decision support systems: an experimental study | 2023 | Students and physicians | 6 clinical scenarios | 7 point scale confidence as well as likelihood of each differential |
| Oskay, A. | Evaluation of thoracic computed tomography interpretation by emergency medicine residents with regards to accuracy and confidence | 2023 | Emergency Medicine | 30 CT scans | 1-10 Confidence |
| Marx, G.; Koens, S.; Von Dem Knesebeck, O.; Scherer, M. | Age and gender differences in diagnostic decision-making of early heart failure: Results of a mixed-methods interview-study using video vignettes | 2022 | General Practice | Video vignettes | 0-100% certainty |
| Albrechtsen, S.S.; Riis, R.G.C.; Amiri, M.; Tanum, G.; Bergdal, O.; Blaabjerg, M.; Simonsen, C.Z.; Kondziella, D. | Impact of MRI on decision-making in ICU patients with disorders of consciousness | 2022 | ICU | Real patient cases in ICU | 5 point likert scale |
| Fawver, B.; Thomas, J.L.; Drew, T.; Mills, M.K.; Auffermann, W.F.; Lohse, K.R.; Williams, A.M. | Seeing isn’t necessarily believing: Misleading contextual information influences perceptual-cognitive bias in radiologists. | 2020 | Radiology | 16 deidentified musculoskeletal radiographic cases | 5 point likert scale |
| Hausmann, D.; Kiesel, V.; Zimmerli, L.; Schlatter, N.; von Gunten, A.; Wattinger, N.; Rosemann, T. | Sensitivity for multimorbidity: The role of diagnostic uncertainty of physicians when evaluating multimorbid video case-based vignettes | 2019 | General Practice / Emergency Medicine | Video vignettes | 0-100% scale confidence in diagnosis |
| Lambe, K.A.; Hevey, D.; Kelly, B.D. | Guided reflection interventions show no effect on diagnostic accuracy in medical students | 2018 | Medical Students | Fictional patient cases | 1-6 scale of confidence in original differential |
| Cairns, A.W.; Bond, R.R.; Finlay, D.D.; Breen, C.; Guldenring, D.; Gaffney, R.; Gallagher, A.G.; Peace, A.J.; Henn, P. | A computer-human interaction model to improve the diagnostic accuracy and clinical decision-making during 12-lead electrocardiogram interpretation | 2016 | GPs and Undergrads | ECG interpretation | Self-rated confidence 1-10 |
| Ben-Assuli, O.; Sagi, D.; Leshno, M.; Ironi, A.; Ziv, A. | Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study | 2015 | Emergency Medicine | Simulated patient scenarios with actors for presenting complaints | 7 point likert scale of confidence in diagnosis |
| Maserejian, N.N.; Lutfey, K.E.; McKinlay, J.B. | Do physicians attend to base rates? prevalence data and statistical discrimination in the diagnosis of coronary heart disease: Physicians and coronary heart disease | 2009 | Primary Care | Vignettes of CHD | 0-100 scale of certainty |
| Abujudeh, H.H.; Kaewlai, R.; McMahon, P.M.; Binder, W.; Novelline, R.A.; Gazelle, G.S.; Thrall, J.H. | Abdominopelvic CT increases diagnostic certainty and guides management decisions: A prospective investigation of 584 patients in a large academic medical center | 2011 | Emergency Medicine | Real patients presenting with abdomen pain | 0-100% certainty |
| van Hout, H.P.J.; Vernooij-Dassen, M.J.; Stalman, W.A.B. | Diagnosing dementia with confidence by GPs | 2007 | General Practice | Observation of dementia patients | 4 point likert scale |
| Benvenuto-Andrade, C.; Dusza, S.W.; Hay, J.L.; Agero, A.L.C.; Halpern, A.C.; Kopf, A.W.; Marghoob, A.A. | Level of confidence in diagnosis: Clinical examination versus dermoscopy examination | 2006 | Dermatology | 20 pairs of clinical and dermoscopic images of lesions | 7 point likert scale of confidence in diagnosis (whether benign or malignant) |
| Dreiseitl, S.; Binder, M. | Do physicians value decision support? A look at the effect of decision support systems on physician opinion | 2005 | Dermatology | 25 dermoscopic lesions | 1-10 scale of benign to malignant, with higher values interpreted as confident? |
| Davis, D.P.; Campbell, C.J.; Poste, J.C.; Ma, G. | The association between operator confidence and accuracy of ultrasonography performed by novice emergency physicians | 2005 | Emergency Medicine | Ultrasound scanning | 1-10 scale of confidence of correct test identification |
| McKinlay, J.B.; Lin, T.; Freund, K.; Moskowitz, M. | The unexpected influence of physician attributes on clinical decisions: Results of an experiment | 2002 | Primary Care | 2 Video vignettes | Certainty adhering to diagnosis (% likelihood for each differential) |
| Friedman, C.; Gatti, G.; Elstein, A.; Franz, T.; Murphy, G.; Wolf, F. | Are clinicians correct when they believe they are correct? Implications for medical decision support | 2001 | Internal Medicine | 36 clinical cases split into 4 equal groups | Confidence in each diagnosis |
| Harvey, C.J.; Halligan, S.; Bartram, C.I.; Hollings, N.; Sahdev, A.; Kingston, K. | Evacuation proctography: A prospective study of diagnostic and therapeutic effects | 1999 | Radiology | Questionnaires after proctography in 50 patient cases | 1-10 confidence in diagnosis |
| Berner, E.S.; Maisiak, R.S. | Influence of case and physician characteristics on perceptions of decision support systems | 1999 | General Practice / Emergency Medicine | Written cases | 1-5 confidence |
| Hillson, S.D.; Connelly, D.P.; Liu, Y. | The Effects of Computer-assisted Electrocardiographic Interpretation on Physicians' Diagnostic Decisions | 1995 | Primary Care | ECG interpretation + vignettes (10) | 1-10 confidence in diagnosis |
| Calman, N.S.; Hyman, R.B.; Licht, W. | Variability in consultation rates and practitioner level of diagnostic certainty | 1992 | GP / Family practice | Observational of consultations | Confidence scored based on physician notes by coders |
| Sklar, D.P.; Hauswald, M.; Johnson, D.R. | Medical problem solving and uncertainty in the emergency department | 1991 | Emergency Medicine | Real patients, filling in questionnaire | Visual analogue scale for each differential |
| Brannon, Laura A; Carson, Kimi L | Nursing expertise and information structure influence medical decision making | 2003 | Nursing | Patient scenarios, manipulated information | 0-100% scale confidence in diagnosis |
| Clayton, Dayna A.; Eguchi, Megan M.; Kerr, Kathleen F.; Miyoshi, Kiyofumi; Brunyé, Tad T.; Drew, Trafton; Weaver, Donald L.; Elmore, Joann G. | Are Pathologists Self-Aware of Their Diagnostic Accuracy? Metacognition and the Diagnostic Process in Pathology | 2023 | Pathology | Diagnosis based on slides for microscopes | 6 point scale confidence in correct diagnosis |
| Friedman, Charles P.; Gatti, Guido G.; Franz, Timothy M.; Murphy, Gwendolyn C.; Wolf, Fredric M.; Heckerling, Paul S.; Fine, Paul L.; Miller, Thomas M.; Elstein, Arthur S. | Do physicians know when their diagnoses are correct?: Implications for decision support and error reduction | 2005 | Internal Medicine | 2-4 page medical synopses diagnosis | Likelihood to seek assistance to reach a diagnosis |
| Garbayo, Luciana S.; Harris, David M.; Fiore, Stephen M.; Robinson, Matthew; Kibble, Jonathan D. | A metacognitive confidence calibration (MCC) tool to help medical students scaffold diagnostic reasoning in decision-making during high-fidelity patient simulations | 2023 | Medical Students | High Fidelity Sim (Cases: Heart Failure, Respiratory Distress, DKA, heat exhaustion) | 7 point likert scale of confidence |
| Hautz, Wolf E; Schubert, Sebastian; Schauber, Stefan K; Kunina\_Habenicht, Olga; Hautz, Stefanie C; Kämmer, Juliane E; Eva, Kevin W | Accuracy of self\_monitoring: does experience, ability or case difficulty matter? | 2019 | Medical Students | 6 clinical scenarios | 10 point scale (0% to 100%) |
| Kämmer, Juliane E.; Schauber, Stefan K.; Hautz, Stefanie C.; Stroben, Fabian; Hautz, Wolf E. | Differential diagnosis checklists reduce diagnostic error differentially: A randomised experiment | 2021 | Medical Students / Emergency Medicien | 6 clinical scenarios | 10 point scale of confidence |
| Kostopoulou, Olga; Russo, J. Edward; Keenan, Greg; Delaney, Brendan C.; Douiri, Abdel | Information Distortion in Physicians’ Diagnostic Judgments | 2012 | Primary Care | 3 clinical scenarios each with 2 competing diagnoses | 21 point likelihood |
| Kourtidis, Ploutarchos; Nurek, Martine; Delaney, Brendan; Kostopoulou, Olga | Influences of early diagnostic suggestions on clinical reasoning | 2022 | Family Medicine | 2 patient scenarios with or without diagnostic suggestions | 10 point visual analogue scale of certainty |
| Krupat, Edward; Wormwood, Jolie; Schwartzstein, Richard M; Richards, Jeremy B | Avoiding premature closure and reaching diagnostic accuracy: some key predictive factors | 2017 | Internal Medicine | 4 complex vignettes | 1-100 scale of certainty |
| Leblanc, Vicki R.; Norman, Geoffrey R.; Brooks, Lee R. | Effect of a Diagnostic Suggestion on Diagnostic Accuracy and Identification of Clinical Features: | 2001 | Medical Students | Scenarios with photographs with clinical features |  |
| Redelmeier, Donald A.; Shafir, Eldar | The Fallacy of a Single Diagnosis | 2023 | Primary Care | Series of vignettes to diagnosis COVID | % likelhiood |
| Trueblood, Jennifer S.; Eichbaum, Quentin; Seegmiller, Adam C.; Stratton, Charles; O'Daniels, Payton; Holmes, William R. | Disentangling prevalence induced biases in medical image decision-making | 2021 | Medical Students / Imaging | Cell scans (cancer identification) |  |
| Yang, Huiqin; Thompson, Carl; Bland, Martin | The effect of clinical experience, judgment task difficulty and time pressure on nurses’ confidence calibration in a high fidelity clinical simulation | 2012 | Nursing | High Fidelity Sim | 0-100 confidence |
| Yang, Huiqin; Thompson, Carl | Nurses’ risk assessment judgements: a confidence calibration study: Nurses’ risk assessment judgements | 2010 | Nursing | Risk assessment vignettes | 0-100 confidence |
| Eva, Wayne Kevin | The influence of differentially processing evidence on diagnostic decision-making | 2001 | Medical Students | Presenting case histories | Probability ratings |
| Tabak, Nili; Bar-Tal, Yoram; Cohen-Mansfield, Jiska | Clinical decision making of experienced and novice nurses | 1996 | Nursing | Two scenarios | 0-100% scale confidence in diagnosis |
| Brezis, Mayer; Orkin-Bedolach, Yael; Fink, Daniel; Kiderman, Alexander | Does Physician's Training Induce Overconfidence That Hampers Disclosing Errors? | 2019 | Cross Disciplines | Survey with clinical vignette of a girl with urinary infection and penicillin allergy | 5 point likert scale |
| Mann, Doug | The Relationship between Diagnostic Accuracy and Confidence in Medical Students. | 1993 | Medical Students / Cardiac | ECG slides - Classification of cardiac dysrhythmias | 11 point scale, 0-100% |
| Schoenherr, Jordan Richard; Waechter, Jason; Millington, Scott J | Subjective awareness of ultrasound expertise development: individual experience as a determinant of overconfidence | 2018 | Cardiology | Cardiac ultrasound case studies | 6 point scale confidence in correct identification |
| Meyer, Ashley ND; Payne, Velma L; Meeks, Derek W; Rao, Radha; Singh, Hardeep | Physicians’ diagnostic accuracy, confidence, and resource requests: a vignette study | 2013 | Internal Medicine | 4 case vignettes | 0-10 confidence in diagnosis (for each) |
| Wood, Greg; Batt, Jeremy; Appelboam, Andrew; Harris, Adrian; Wilson, Mark R. | Exploring the Impact of Expertise, Clinical History, and Visual Search on Electrocardiogram Interpretation\*\* | 2014 | ED | ECG traces and eye tracking | 1-10 confidence in diagnosis |
| Bergl, P. A.; Shukla, N.; Shah, J.; Khan, M.; Patel, J. J.; Nanchal, R. S. | Factors influencing diagnostic accuracy among intensive care unit clinicians – an observational study\*\* | 2024 | ICU | Surveys during ICU | 5 point likert scale |
| Frey, J.; Braun, L. T.; Handgriff, L.; Kendziora, B.; Fischer, M. R.; Reincke, M.; Zwaan, L.; Schmidmaier, R. | Insights into diagnostic errors in endocrinology: a prospective, case-based, international study\*\* | 2023 | Endocrinology | 5 patient cases | 1-10 confidence in diagnosis |
| van Sassen, C.; Mamede, S.; Bos, M.; van den Broek, W.; Bindels, P.; Zwaan, L. | Do malpractice claim clinical case vignettes enhance diagnostic accuracy and acceptance in clinical reasoning education during GP training?\*\* | 2023 | General Practice | Cases with and without malpractice claim information | 0-100 confidence |
| Gupta, A. B.; Greene, M. T.; Fowler, K. E.; Chopra, V. I. | Associations Between Hospitalist Shift Busyness, Diagnostic Confidence, and Resource Utilization: A Pilot Study\*\* | 2023 | Doctors | Questionnaire during shift | 1-10 Confidence |
| Kuhn, J.; Mamede, S.; van den Berg, P.; Zwaan, L.; van Peet, P.; Bindels, P.; van Gog, T. | Learning deliberate reflection in medical diagnosis: does learning-by-teaching help?\*\* | 2023 | General Practice | 10 written cases | 1-9 confidence |
| Staal, J.; Katarya, K.; Speelman, M.; Brand, R.; Alsma, J.; Sloane, J.; Van den Broek, W. W.; Zwaan, L. | Impact of performance and information feedback on medical interns' confidence–accuracy calibration\*\* | 2023 | Medical Students | X-ray interpretation | 0-10 confidence in diagnosis |
| Keene, T.; Pammer, K.; Lord, B.; Shipp, C. | Dispatch information affects diagnosis in paramedics: an experimental study of applied dual-process theory\*\* | 2022 | Paramedics | Vignettes in two parts with an intuitive impression and then diagnosis, with or wtihout secondary task distraction | 4 point scale |
| Tio, R. A.; Filho, M. A. C.; de Menezes Mota, M. F.; Santanchè, A.; Mamede, S. | The Effect of Information Presentation Order on Residents’ Diagnostic Accuracy of Online Simulated Patients With Chest Pain\*\* | 2022 | Cardiology | 12 clinical cases presented in 2 diagnostic rounds (history and EKG) | 0-100 confidence |
| Kuhn, J.; van den Berg, P.; Mamede, S.; Zwaan, L.; Bindels, P.; van Gog, T. | Improving medical residents’ self-assessment of their diagnostic accuracy: does feedback help?\*\* | 2022 | General Practice | 12 cases | 1-9 confidence |
| Katz, I.; O'Brien, B.; Clark, S.; Thompson, C. T.; Schapiro, B.; Azzi, A.; Lilleyman, A.; Boyle, T.; Espartero, L. J. L.; Yamada, M.; Prow, T. W. | Assessment of a Diagnostic Classification System for Management of Lesions to Exclude Melanoma\*\* | 2021 | Pathology / Dermatology | 217 Lesions prepared and stained from patients | 1-5 confidence |
| Staal, J.; Alsma, J.; Mamede, S.; Olson, A. P. J.; Prins-van Gilst, G.; Geerlings, S. E.; Plesac, M.; Sundberg, M. A.; Frens, M. A.; Schmidt, H. G.; Van den Broek, W. W.; Zwaan, L. | The relationship between time to diagnose and diagnostic accuracy among internal medicine residents: a randomized experiment\*\* | 2021 | Internal Medicine | 8 clinical case | 0-100% scale confidence that diagnosis was correct |
| Thorlacius-Ussing, G.; Bruun, M.; Gjerum, L.; Frederiksen, K. S.; Rhodius-Meester, H. F. M.; Van Der Flier, W. M.; Waldemar, G.; Hasselbalch, S. G.; Nobili, F. | Comparing a Single Clinician Versus a Multidisciplinary Consensus Conference Approach for Dementia Diagnostics\*\* | 2021 | Neurology | Real patient evaluations | 0-100 Visual analogue scale |
| Chen, Y.; Nagendran, M.; Kilic, Y.; Cavlan, D.; Feather, A.; Westwood, M.; Rowland, E.; Gutteridge, C.; Lambiase, P. D. | The diagnostic certainty levels of junior clinicians: A retrospective cohort study\*\* | 2021 | Emergency Medicine | Real patient cases deindentified | Qualitative labels translated into % |
| Li, S.; Zheng, J.; Lajoie, S. P. | The relationship between cognitive engagement and students’ performance in a simulation-based training environment: an information-processing perspective\*\* | 2020 | Medical Students | Two patient cases shown | 0-100% scale confidence in diagnosis |
| Chartan, C.; Singh, H.; Krishnamurthy, P.; Sur, M.; Meyer, A.; Lutfi, R.; Stark, J.; Thammasitboon, S. | Isolating red flags to enhance diagnosis (I-RED): An experimental vignette study\*\* | 2019 | Paediatric residents | Paediatric cases | 1-10 Confidence |
| Cleary, T. J.; Konopasky, A.; La Rochelle, J. S.; Neubauer, B. E.; Durning, S. J.; Artino, A. R. | First-year medical students’ calibration bias and accuracy across clinical reasoning activities\*\* | 2019 | Medical Students | Some of kind of virtual patient sim | Estimations of performance |
| Costa Filho, G. B.; Moura, A. S.; Brandão, P. R.; Schmidt, H. G.; Mamede, S. | Effects of deliberate reflection on diagnostic accuracy, confidence and diagnostic calibration in dermatology\*\* | 2019 | Medical Students / dermatology | 12 dermatological images | 0-100% scale confidence in diagnosis |
| Nederhand, M. L.; Tabbers, H. K.; Splinter, T. A. W.; Rikers, R. M. J. P. | The Effect of Performance Standards and Medical Experience on Diagnostic Calibration Accuracy\*\* | 2018 | General Medicine | 6 clinical cases | Confidence in diagnosis (1-10) |
| Pusic, M. V.; Chiaramonte, R.; Gladding, S.; Andrews, J. S.; Pecaric, M. R.; Boutis, K. | Accuracy of self-monitoring during learning of radiograph interpretation\*\* | 2015 | Radiology / medical students | Ankle radiographs | Qualitative labels |
| Hautz, W. E.; Kämmer, J. E.; Schauber, S. K.; Spies, C. D.; Gaissmaier, W. | Diagnostic performance by medical students working individually or in teams\*\* | 2015 | Medical Students | 6 simulated cases of respiratory distress | 1-10 Confidence |
| Soares, W. E.; Price, L. L.; Prast, B.; Tarbox, E.; Mader, T. J.; Blanchard, R. | Accuracy screening for ST elevation myocardial infarction in a task-switching simulation\*\* | 2019 | Emergency Medicine | ECG interpretation | 1-5 confidence |
| Blissett, S.; Sibbald, M.; Kok, E.; van Merrienboer, J. | Optimizing self-regulation of performance: is mental effort a cue? \*\* | 2018 | Internal Medicine | ECG interpretation | 0-100% certainty |
| Adderley, U. J.; Thompson, C. | Confidence and clinical judgement in community nurses managing venous leg ulceration – A judgement analysis\*\* | 2017 | Nursing | 110 (!) clinical scenarios | 1-10 confidence in diagnosis |
| Feyzi-Behnagh, R.; Azevedo, R.; Legowski, E.; Reitmeyer, K.; Tseytlin, E.; Crowley, R. S. | Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system\*\* | 2014 | Pathology / Dermatology | Dermatoligical slides | 6 point scale confidence in correct diagnosis |
| Hageman, M. G. J. S.; Bossen, J. K. J.; King, J. D.; Ring, D. | Surgeon confidence in an outpatient setting\*\* | 2013 | Surgery | Real patients visiting surgery | 5 point likert scale |
| Crowley, R. S.; Legowski, E.; Medvedeva, O.; Reitmeyer, K.; Tseytlin, E.; Castine, M.; Jukic, D.; Mello-Thoms, C. | Automated detection of heuristics and biases among pathologists in a computer-based system\*\* | 2013 | Pathology / Dermatology | Dermatoligical slides | Scale from -1 to +1 |
| Yang, H.; Thompson, C.; Bland, M. | Effect of improving the realism of simulated clinical judgement tasks on nurses' overconfidence and underconfidence: Evidence from a comparative confidence calibration analysis\*\* | 2012 | Nursing | Both paper and high fidelity sim scenarios | 0-100 confidence |
| Gruppen, L; Wolf, F; Billi, J | Information Gathering and Integration as Sources of Error in Diagnostic Decision Making\*\* | 1991 | Primary Care | Vignettes deciding between two diagnostic alternatives | Probability correct |

Studies marked with \*\* next to their title were included via citation tracking