

Confidence and Certainty in Medical Diagnoses within Acute Healthcare: A Scoping Review

Sriraj Aiyer¹, Helen Higham^{2,3} & Nick Yeung¹

1. Department of Experimental Psychology, University of Oxford
2. Oxford University Hospitals NHS Foundation Trust
3. Nuffield Department of Clinical Neurosciences, University of Oxford

1 **Abstract**

2

3 Objective

4 Overconfidence is an important source of medical error. This review analyses experimental
5 studies of confidence in medical diagnosis to identify factors affecting clinicians' confidence
6 in their diagnoses, and how confidence impacts patient care.

7

8 Method

9 A scoping review of medical and psychological literature was conducted. Articles were
10 categorised according to methodology and clinical speciality. Findings were analysed
11 thematically. Our review methodology adheres to the JBI's PRISMA-ScR Checklist for Scoping
12 Reviews.

13

14 Data Sources

15 We searched SCOPUS, MEDLINE, PsycINFO and Global Health. We then performed citation
16 tracking within these papers' references to identify additional articles.

17

18 Eligibility criteria

19 Papers were included if they reported quantitative results from an empirical study in which
20 participants reported their confidence or certainty during a diagnostic decision. Studies
21 comprised several medical subdisciplines.

22

23 Results

24 77 articles met the inclusion criteria. Across these articles, confidence was not found to be
25 well-calibrated to true diagnostic accuracy regardless of clinician experience. We organised
26 articles under two main themes: the determinants of confidence and the uses of confidence
27 during the patient's care pathway. Confidence is found to be affected by several factors
28 including case complexity, early diagnostic differentials, and the healthcare environment.
29 Factors that affect confidence, but not accuracy, demonstrate how the two can become
30 decoupled, resulting in overconfidence/underconfidence. Confidence is found to affect
31 patient testing, medication administration and referral rates, among other clinical actions.

32

33 Conclusions

34 Improving the calibration of confidence should be a priority for medical education and
35 clinical practice (e.g., via decision aids). We propose a theoretical model of factors that
36 affect diagnostic confidence/certainty. Such a model can inform future work on how
37 appropriate diagnostic confidence can be prompted and communicated amongst clinicians.

38

39 Word Count: 275

40 WHAT IS ALREADY KNOWN ON THIS TOPIC

41 There is extensive evidence of diagnostic error in most healthcare specialities, leading to
42 suggestions that cognitive biases, such as overconfidence, are causally linked with these
43 errors.

44

45 WHAT THIS STUDY ADDS

46 This review identifies that diagnostic confidence and accuracy have separate contributing
47 factors related to the patient's condition, the clinician making the diagnosis, and the
48 healthcare context. It also identifies pathways by which confidence, and misplaced
49 confidence, can affect diagnosis and patient care.

50

51 HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

52 Firstly, there is little evidence that clinicians' confidence is aligned to their diagnostic
53 accuracy, even when using certain cognitive interventions or aids, across several medical
54 disciplines. Secondly, confidence is predictive of actions in many parts of the patient care
55 process, such as ordering investigations, referrals to specialists or prescribing, which may be
56 suboptimal if confidence is miscalibrated. The proposed conceptual model of confidence
57 during the patient care process synthesises current findings, identifies both individual and
58 contextual factors that inform confidence, and recommends how future research can focus
59 on underexplored areas, particularly on group decisions, individual differences in
60 confidence, and the link between information seeking and confidence.

61

62 INTRODUCTION

63

64 Accurate medical diagnosis is crucial to safe, high quality patient care. Research on diagnosis
65 within acute healthcare has been grounded in the incidence of errors. A report from the US
66 Institute of Medicine¹ concluded that most patients will experience a diagnostic error within
67 their lifetime. Around 32% of clinical errors arise during clinician assessment, particularly
68 clinicians' failure to weigh up competing diagnoses². Diagnostic errors have substantial
69 downstream consequences: Unnecessary treatment (or 'overtreatment') was estimated to
70 cost the US healthcare system \$158-\$226 billion in 2011³. Diagnostic errors lead to longer
71 hospital stays and increased patient mortality⁴.

72

73 Heuristics are commonly used in diagnostic decisions, which are necessarily based on
74 incomplete, imperfect information and made under time pressure. For example, making a
75 diagnosis may involve considering a hypothesis as likely because the displayed symptoms
76 correspond with a prototypical case of a particular condition⁵, or match a recent experience
77 of a patient with a specific condition⁶. Although these heuristics are often effective,
78 associated cognitive biases are thought to be linked to diagnostic error. One cognitive bias
79 that has been posited as a contributor to diagnostic error is overconfidence⁷, which may
80 cause an individual clinician to fail to consider alternative diagnoses or lead a clinical team
81 to be overly swayed by one individual's opinion. The accuracy of diagnostic decisions, which
82 is considered within existing models of diagnosis^{8,9}, is only inferred indirectly from a
83 clinician's subjective confidence. Given that the ground truth of a patient's condition (and as
84 such, the accuracy of clinicians' diagnosis) is often unknowable, and confidence is

85 communicated between clinicians as a marker of accuracy, diagnostic confidence is an
86 important area of study.

87

88 Confidence is the subjective assessment of a decision's quality or accuracy¹⁰. We refer to
89 confidence as 'calibrated' if it closely predicts objective accuracy (i.e., such that the
90 individual is neither overconfident nor underconfident, Figure 1). In experimental studies,
91 confidence can exhibit impressive calibration¹¹, reflecting people's ability to evaluate the
92 quality of evidence on which they base their decisions¹². But calibration is rarely perfect
93 because confidence also depends on factors that do not directly correlate with accuracy,
94 such as the time spent deliberating and the total amount of evidence considered
95 (independent of the quality or consistency of this evidence)^{13,14}, as well as the mood¹⁵,
96 personality¹⁶, gender¹⁷ and status¹⁸ of the decision maker. The resulting under- and
97 overconfidence matters: overconfident decision makers leap to premature conclusions and
98 ignore useful advice, while underconfident decision makers waste time collecting evidence
99 that will not improve their decisions¹⁹. Effective decision making in groups likewise depends
100 on team members sharing calibrated information about their uncertainty: Confident team
101 members tend to be listened to more, which can lead others astray if they are
102 overconfident²⁰. Conversely, underconfident team members may be ignored or may fail to
103 share potentially useful information²¹.

104

105 [INSERT FIGURE 1]

106

107 These features of confidence highlight its potential importance in healthcare, as
108 overconfidence can lead to insufficient consideration of diagnostic alternatives and
109 inadequate care²². In the absence of objective feedback, confidence can be used as a
110 marker of how likely someone is to be correct²³. In medicine, a lack of clearly communicated
111 feedback can cause clinicians to proceed as if they have received positive feedback. This
112 means that they may not adequately update their internal model of the patient and then
113 increase their confidence inappropriately, whether working individually or in teams²⁴.

114

115 This scoping review collates and synthesises existing work studying diagnosis as a cognitive
116 process. Although past work has reviewed literature on confidence calibration in mental
117 health diagnoses²⁵, this is the first scoping review, to our knowledge, with such a remit to
118 include studies of confidence across medical subdisciplines, given its broad importance
119 across medicine. The use of scoping review is suitable given this remit. We aimed to identify
120 key determinants of confidence and its role in medical decision making. In common with
121 practice in the reviewed articles, we use "confidence" and "certainty" interchangeably,
122 while noting that the psychological literature discusses whether they are subtly different
123 concepts²⁶. Our full research questions can be found in Box 1.

124

125

126

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

128

129 Primary questions:

130

- 131 - How calibrated are confidence/certainty judgements made during diagnostic
132 decisions by clinicians relative to their actual accuracy?
133 - How are confidence/certainty judgements utilised within the wider diagnostic
134 decision process?
135

136 Subsidiary questions:

- 137
- 138 - What types of empirical procedures/tasks are used to study confidence and
139 certainty in diagnostic decisions? Do they come to different conclusions?
 - 140 - What are the prevalent ways in which diagnostic confidence and certainty are
141 measured and operationalised as variables?
 - 142 - What strategies, tools or frameworks have been used to prompt better calibration of
143 confidence and certainty?
 - 144 - What are the discrepancies between the concepts/research questions studied in the
145 context of confidence and certainty in the cognitive psychology literature and the
146 medical diagnosis literature?
 - 147 - What areas of research are still underexplored within the context of medical
148 diagnosis?
149
-

150

151 **METHODS**

152

153 **Search Strategy**

154

155 Our review protocol was preregistered (<https://osf.io/wz5se>). We conducted a scoping
156 review of empirical studies on confidence and certainty in medical diagnosis using JBI's
157 PRISMA-ScR Checklist for Scoping Reviews²⁷. The search strategy was designed in
158 cooperation with a subject specialist librarian at the University of Oxford's Bodleian Library.
159 The search string comprised keywords capturing the intersection of four elements:
160 confidence/certainty, medical diagnoses, decision making and a study population of medical
161 staff/students (i.e., clinicians, physicians, doctors and medics). The full search terms can be
162 found in Box 2. SCOPUS, MEDLINE, PsycINFO and Global Health databases were searched
163 during February 2024. Finally, we hand-searched the citations of the included articles from
164 these databases for further relevant articles via backward and forward scanning^{28,29}.
165

166

167

168

168 **BOX 2: Search Details**

169

170 **Search Terms**

171

172 (clinicians OR physicians OR doctors OR medics)

173 AND

174 (confiden* OR uncertain* OR certain*)

175 AND

176 (diagnosis AND medical)

177 AND

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

179

180

181 **Study Selection**

182

183 • inclusion criteria: (1) original empirical studies with quantitative results, (2) written
184 in the English language, (3) experimental paradigm uses medical diagnostic
185 decisions, (4) confidence or certainty is measured as a dependent variable (5) any
186 medical discipline.

187 • Exclusion criteria: editorials, review papers and opinion papers (dissertations with
188 original empirical work were included)

189 • As we are specifically interested in the diagnosis of physical illnesses and injuries, our
190 population of interest for studies was medical students, clinicians or medics.

191

192 Identified articles were uploaded onto Rayyan (<https://rayyan.ai/>) to detect duplicate
193 papers for manual checking and removal. This is the only part of the review process in which
194 an automation tool is used. Data from the papers was collected using manual review.

195

196 **Research Synthesis**

197

198 Papers selected for review were first categorised by their broad research methodology (e.g.,
199 patient vignettes, in situ questionnaires, etc.) and their medical population of study (e.g.,
200 medical students, general practitioners/hospital physicians etc.). We also extracted
201 information such as sample size and the variables manipulated/measured by the
202 researchers. Research characteristics to be extracted were derived iteratively between the
203 authors with regular meetings and can be found on OSF (<https://osf.io/4g6s3/>). We provide
204 a full list of extracted data in the Supplemental Materials (Table S1). We reviewed the
205 experimental procedures to extract their key manipulations and independent variables (e.g.,
206 case complexity, use of a cognitive intervention, level of medical experience). We also
207 extracted dependent variables as they pertain to confidence/certainty and, where relevant,
208 recording of both diagnostic differentials and information seeking. Each of the paper's key
209 findings were briefly summarised and then the authors categorised findings based on our
210 two primary research questions and on recurring research themes. As this is a scoping
211 review that aims to map the existing research landscape, we do not make any assessments
212 of study quality.

213

214

215

216 **RESULTS**

217

218 **Findings of Scoping Review**

219

220 Initial search returned 3,332 articles. 48 eligible articles met the inclusion criteria. 439
221 further articles were retrieved for review from the included articles' citations. After applying
222 both exclusions of duplicates and our inclusion criteria, 29 further articles were identified.

223 Altogether, 77 articles were reviewed (see Figure 2 for PRISMA diagram). The full set of
224 papers can be found in Table S2 (Supplemental Materials).

225
226 [INSERT FIGURE 2]

227 228 **Study Characteristics**

229
230 Table 1 summarises study characteristics. 35 of the 77 studies (45%) were published since
231 2019, indicating a recent surge of research interest and the timeliness of the present review.
232 The studies appeared in 59 different publications, including both medical and psychological
233 journals, with medical education journals being most common (18 studies). Clinical areas
234 most represented were Primary Care/General Practice, Emergency Medicine and Nursing.

235 236 **What types of empirical procedures/tasks are used to study confidence and** 237 **certainty in diagnostic decisions?**

238
239 Study designs were split roughly evenly between focusing on how confidence varies across
240 individuals (34 studies) and how confidence varies according to features of the patient case
241 (30 studies), with remaining (13 studies) studying both. Thirty-four studies (44%) looked at
242 the level of medical experience or training's effect on confidence, either measured as a
243 dependent variable or by recruiting participants in 'novice' versus 'experienced' group.
244 Nineteen studies (25%) manipulated patient case complexity or difficulty. Finally, ten
245 studies (13%) investigated how diagnostic confidence varies with the information presented
246 or the opportunity to seek information.

247
248 Most studies (57%) used clinical patient text vignettes. For vignettes, there is an established
249 ground truth in each case (which may not be possible for in situ studies involving real
250 patients) to compare participants' confidence to their objective accuracy in order to gauge
251 calibration. Because vignettes are quick and simple to administer, participants can complete
252 several diagnoses during a single study such that both their confidence and accuracy can be
253 averaged across cases. Other experimental methodologies include the use of imaging (e.g.,
254 ECG, X-Rays, MRI) for diagnosis, high-fidelity simulations (either using extended reality tools
255 or a patient mannequin), or questionnaires administered in situ to measure confidence
256 during real patient cases (Table 1). The preponderance of vignette studies is noteworthy
257 given one finding that nurses were both less accurate and less confident in a high-fidelity
258 simulation compared to a paper-based vignette³¹, suggesting the need for caution when
259 generalising experimental findings to how medical professionals behave in everyday
260 practice.

261 262 **What are the prevalent ways in which diagnostic confidence and certainty** 263 **are measured and operationalised as variables?**

264
265 Studies varied in how confidence and diagnostic accuracy were assessed. Most studies used
266 a self-reported numerical scale for confidence (usually 1-10 or 1-100), whereas verbal
267 expressions of confidence (e.g., "not sure" to "certain") were used in three studies, and
268 visual analogue scales were used in four studies. The use of numerical scales is common

269 within cognitive psychology, where measured confidence values predict other behavioural
 270 indices of uncertainty, such as the tendency to seek further information or to opt out of
 271 making a decision³². Twenty-four studies (31%) allowed participants to input multiple
 272 diagnostic differentials rather than a single diagnosis. Confidence is then either measured
 273 for each differential or in the set of differentials as a whole. Confidence tended to be
 274 measured at the point of diagnosis or shortly afterwards. Only five studies recorded
 275 confidence before a final diagnosis was provided to investigate how confidence changes
 276 over the course of the diagnostic process, rather than at a single decision point.

277
 278 In terms of accuracy, most studies prompt clinicians for a single diagnosis that is marked as
 279 correct or incorrect. However, clinicians may consider multiple possible diagnoses in their
 280 everyday practice. Hence, 24 studies (31%) allowed participants to record multiple
 281 differentials. This complicates scoring accuracy and confidence: If accuracy is
 282 operationalised as whether a correct diagnosis is included in this set of multiple
 283 differentials, clinicians are more likely to be correct with more differentials, and it may
 284 remain unclear how clinicians weigh up competing differentials. Hence, the operationalised
 285 calibration of confidence judgements is heavily contingent on how diagnoses are recorded.
 286 Twenty studies (26%) did not measure or report accuracy, predominantly because they
 287 utilise paradigms/scenarios in which there was no established ground truth against which to
 288 evaluate accuracy, for instance in involving real patient cases. Accordingly, calibration is not
 289 measured for these studies either, given that a measure of calibration requires an objective
 290 standard of performance to compare confidence against. Studies that measured accuracy
 291 either used pre-determined criteria for correct diagnoses or expert clinicians' independent
 292 assessments of accuracy.

Publication Year		Subdiscipline / Population	
1991-2000	10	Primary Care / General Practice	26
2001-2010	10	Medical Students	13
2011-2020	29	Emergency Medicine	10
2021-	28	Nursing	6
		Pathology	4
Study Environment/Context		Radiology	4
Text Vignette	44	Other	14
Imaging Interpretation (e.g. ECG)	18		
In Situ Questionnaires/Surveys	13	Study Population Sample Size	
High-Fidelity Simulation	2	< 100	42
		101-200	20
Participant Experience Levels		201-300	9
Participants with Similar Experience	44	> 300	6
Multiple Experience Groups Recruited	28		
Experience Used for Exploratory Analyses	5	Total	77

293
 294 **TABLE 1: Characteristics of Included Studies, including year of publication, study**
 295 **environment used and medical population (recruiting single or multiple levels of**

296 participant experience, medical subdiscipline, sample size). A full list of all included papers
297 can be found in the Supplemental Materials (Table S1).

298
299

300 **Emerging Research Themes**

301

302 *How calibrated are confidence/certainty judgements made during diagnostic decisions by*
303 *clinicians relative to their actual accuracy?*

304

305 Of the included papers, 33 studies analysed the calibration of the participants' diagnostic
306 confidence. Of the remaining articles, 20 studies did not include a measure of accuracy,
307 whilst 24 studies had measures of both confidence and accuracy, but did not analyse
308 confidence calibration.

309

310 Calibration is assessed by comparing confidence ratings with objective diagnostic accuracy:
311 When clinicians rate 100% (or 50% or 60%, etc.) certainty in their diagnosis, are they in fact
312 correct 100% (or 50%, 60%, etc.) of the time? Calibration is then an indirect measure that is
313 calculated by comparing two other observed measures: confidence and accuracy.

314 Calibration can be defined in a number of ways, with some studies reporting/analysing
315 multiple measures. One is to calculate the difference between confidence and accuracy
316 when both are defined as percentages (used by 13 studies). Another measure (11 studies),
317 commonly referred to as 'resolution', involves calculating the difference in confidence when
318 clinicians are objectively correct versus incorrect. The third common measure is to use a
319 regression/correlation analysis to determine if accuracy and confidence are closely related
320 to each other, either within individuals (across multiple cases) or across individuals (13
321 studies).

322

323 In our study sample, there was limited evidence of calibrated confidence judgements, with
324 some studies reporting underconfidence^{30,33,34} and others overconfidence³⁵⁻³⁷. To examine
325 these findings in more detail, we considered factors that impact and promote calibration in
326 diagnoses.

327

328 *The Impact of Experience on Calibration*

329

330 The first major theme of interest is whether calibration improves with experience. This was
331 not always observed in the results^{38,39}. However, experienced clinicians seem better able to
332 identify complex cases and adjust their confidence accordingly^{40,41}. Looking at the link
333 between calibration and experience alone may be too simplistic, and there are other
334 aspects of experience that influence diagnoses. Experienced clinicians were found to be less
335 likely to 'distort' neutral information to be in support of their reported diagnoses, indicating
336 a lower tendency toward confirmation bias⁴². Past work has also suggested a distinction
337 between experience (operationalised as years of experience or role seniority) and
338 knowledge (measured using standardised tests of medical knowledge). In medical students,
339 the calibration of confidence judgements were found to improve with years of education
340 but not with medical knowledge⁴³. Information 'distortion' was found to affect novice
341 clinicians more⁴² and lower knowledge was found to be related to higher susceptibility to
342 irrelevant, distracting features of a patient⁴⁴. However, the latter study from Mamede et al

343 (2024) found that medical knowledge on the part of resident physicians was not directly
344 associated with calibration.

345
346 *The Impact of Contextual Factors on Calibration*
347

348 The second major theme relates to contextual and environmental factors. Studies have
349 found that calibration is affected by case complexity or difficulty⁴⁵⁻⁴⁷. When confidence
350 judgements are not sensitive to the difficulty or complexity, confidence stays fairly constant
351 for difficult cases whilst accuracy decreases, leading to increased overconfidence (and
352 decreased calibration)⁴⁵. In past studies, complexity is manipulated by either presenting
353 patient cases with more comorbid conditions⁴⁶ or by showing conflicting information about
354 the patient to indicate multiple possible conditions³⁸. Calibration can be improved by the
355 presence of feedback during a training period^{48,49}.

356
357 Contextual factors in the medical environment can also affect overall levels of confidence
358 (rather than calibration), as found using naturalistic paradigms. For example, clinicians may
359 be constantly interrupted by other tasks⁵⁰, especially during busier shifts where they have to
360 manage more patients⁵¹ and may not be present for the sharing of information during
361 handovers⁵². Studies that simulated these situations found they resulted in lower diagnostic
362 confidence. However, such studies cannot assess the effect of contextual factors on
363 confidence calibration with diagnostic accuracy because they were conducted in-situ,
364 meaning that researchers do not yet have a ground truth of the patient's condition. At this
365 stage, we can only determine how these contextual factors affect confidence, rather than
366 calibration.

367
368
369

370 **BOX 3: Papers on Imaging and Decision Support Systems**
371

372 A subset of seven papers found evidence for an increase in confidence when providing
373 clinicians with specialised imaging for a patient to assist diagnoses, be they MRI scans^{53,54},
374 CT scans⁵⁵, evacuation proctography⁵⁶ or photos of wounds⁵⁷. Another subset of nine papers
375 used various forms of computer-aided decision support systems with the goal of increasing
376 confidence⁵⁸⁻⁶¹. The remaining papers did not utilise any imaging or decision support
377 systems. Hillson, Conelly & Liu (1995) found that the adoption of diagnoses that were
378 recommended by a computer-aided decision support system was not associated with an
379 increase in confidence. Neugebauer, Ebert & Vogelmann (2020) did find evidence for such
380 an association, however, with use of decision support leading to both increased confidence
381 and increased diagnostic accuracy. On the other hand, both Berner & Maisiak (1999) and
382 Dreiseitl & Binder (2005) found that usage of decision support recommendations were
383 associated with lower confidence when compared with decisions in which such
384 recommendations were not utilised. Taken as a whole, whilst useful imaging increases
385 confidence when available to clinicians, the efficacy of decision support systems at
386 increasing confidence is likely dependent on other factors that require future work to
387 elucidate.

388
389

390 *What strategies, tools or frameworks have been used to prompt better calibration of*
391 *confidence and certainty?*

392

393 The process of generating diagnostic differentials has been subject to experimental
394 manipulations and interventions (such as early diagnostic suggestions) to investigate their
395 effect on accuracy and confidence. This work is applicable, for instance, to understanding
396 how a clinician transitions care of a patient to another clinician and gives a handover of
397 relevant information. A general theme of this work is that there is a tendency toward higher
398 weighting of early information. Early diagnostic suggestions have been found to be highly
399 influential on subsequent decisions whereby clinicians find these suggestions difficult to
400 ignore and have more confidence in them^{62,63}. This also affects the breadth of differentials
401 considered, with fewer differentials considered when provided with early suggestions⁶⁴ and
402 underweighting of differentials considered later in the diagnostic process⁶⁵. Interventions
403 aimed at mitigating this tendency by asking clinicians to explicitly consider alternatives,
404 increased their accuracy and calibration⁶⁶, or prompting the consideration of the patient's
405 'red flags' in diagnoses, which increased confidence on simpler cases but not accuracy⁶⁷.
406 These interventions seem to require explicit instructions: Simply asking clinicians to reflect
407 on their decision without guidance^{68,69} or participate in an educational training course^{70,71}
408 does not seem to improve diagnostic accuracy and calibration.

409

410 Studies have also investigated how confidence is affected by the manner in which
411 information is presented to clinicians. Higher confidence was found when clinicians were
412 presented with additional patient information, even when this did not carry diagnostic
413 value⁷², and when given all available patient information rather than having to gather
414 information themselves⁷³. Clinicians were also found to be more confident and more
415 accurate when presented with an Electronic Health Record of the patient alongside other
416 information⁷⁴ and when presented with the patient history first rather than later⁷⁵. This
417 finding indicates that complete patient history available early on has a positive impact on
418 confidence. However, erroneous patient history cues both novice and experienced clinicians
419 towards incorrect diagnoses whilst confidence remains relatively high, resulting in
420 overconfidence⁷⁶.

421

422 *How are confidence/certainty judgements utilised within the wider diagnostic decision*
423 *process?*

424

425 More naturalistic studies allow investigation of the role of confidence within the wider
426 diagnostic process, especially where healthcare involves transitions of care between
427 multiple clinicians and departments. Past work has attempted to establish a link between
428 confidence and further seeking of patient information and tests, with mixed results. US
429 hospitalists (medical staff who provide care for patients specifically within US hospitals) with
430 lower confidence were found to order more tests⁵¹ whilst pathologists who were better
431 calibrated (i.e., who tended to report confidence judgements that were closer to true
432 accuracy) were more likely to request further tests when they were unsure⁴¹. Confidence
433 has also been linked to prescribing medication, though overtreatment with unnecessary
434 medications has been linked to both underconfidence⁷⁷ and overconfidence³⁷. Higher
435 confidence is also linked to more referrals to specialists in other departments⁷⁸ and to a
436 lower willingness to admit mistakes³⁴. One study found that whilst experienced clinicians

437 were not more accurate in their initial diagnoses, they were more willing to change
438 diagnoses and request more information⁷⁹. Lower confidence has been found to result in
439 less specific diagnoses for patients in situ⁸⁰. Although psychology research on confidence
440 has examined its role within groups (as discussed in the Introduction), only one article
441 looked at confidence in group decisions in medicine. It found that a multidisciplinary panel
442 was more confident and better calibrated than a single clinician⁸¹.

443

444 *Conceptual Model for Diagnostic Decisions*

445

446 We synthesised the reviewed findings into a theoretical model (Figure 3) that illustrates
447 how various factors distinctly impact diagnostic confidence and accuracy. This model aims
448 to clarify existing research and identify directions for future work, by first building on
449 existing characterisations of the diagnostic decision process^{8,9}. The model starts by mapping
450 out the stages of the diagnostic process (Figure 3, bottom panel). Based on initial patient
451 presentation, clinicians gather and interpret patient information (e.g. history, examinations,
452 tests) to inform their diagnosis. Clinicians' confidence in their diagnosis guides their
453 judgment on when they have enough information to begin treatment versus whether
454 further tests or additional information are needed. Once a diagnosis is reached, this guides
455 patient treatment and care, the success of which is evident in the outcome for the patient,
456 reflecting diagnostic accuracy.

457

458 The middle panel of Figure 3 characterises the cognitive processes of the clinician that
459 determine the accuracy of the diagnosis and confidence with which the diagnosis is made.
460 Critically, many factors have dissociable effects on accuracy vs. confidence. Diagnostic
461 accuracy depends more on the level of medical knowledge and the quality of information
462 gathering and interpretation; confidence depends more on years of medical experience and
463 the quantity of information gathered⁴²⁻⁴⁴. Knowledge is improved through feedback on how
464 a patient case was handled and its outcome, which in turn improves future diagnostic
465 accuracy (though this feedback loop's impact on later confidence is yet to be explored).

466

467 The top panel of Figure 3 highlights factors pertaining to the medical environment/context.
468 Separately from the patient case, confidence is reduced by time pressures³¹, interruptions
469 to work⁵⁰, busy shifts⁵¹ and complex patient cases⁴⁵⁻⁴⁷ (either due to conflicting information
470 or comorbidities).

471

472 Figure 3 highlights three primary directions for future research. First, whereas research to
473 date has focused on diagnosis by individual clinicians, we recommend future work also
474 studies diagnoses in groups, given that diagnoses are often made by teams rather than
475 individuals, particularly in secondary care settings. This is especially pertinent given the
476 social influence that experience/seniority can have within a group: junior clinicians may be
477 less likely to speak up about potential errors⁸² or share valuable case information that only
478 they currently know⁸³ in the presence of more experienced clinicians⁸². Second, future work
479 should study individual differences across clinicians to characterise how personality and
480 trait-level factors such as personality¹⁶, gender¹⁷ and status¹⁸ may impact clinicians'
481 confidence in their diagnoses. Finally, we recommend future work investigate the
482 interaction between the information seeking and confidence formation. Past work has
483 tended to frame information seeking as a further action after diagnosis, rather than an

484 inherent part of the process that forms the diagnosis in the first place. Future work should
485 prioritise examining how to prompt appropriate information seeking (i.e., neither
486 overtesting nor undertesting) via educational tools or cognitive interventions.

487 [INSERT FIGURE 3]

488

489 **DISCUSSION**

490

491 The present work comprehensively maps out the literature on confidence in medical
492 diagnoses, thus extending previous work exploring how cognitive biases contribute to
493 medical error⁸⁴⁻⁸⁶ and characterising medical uncertainty more broadly^{87,88}. This scoping
494 review shows the importance of, and recent surge of interest in, diagnostic confidence.
495 Although confidence has previously been linked to diagnostic error⁷, full understanding will
496 benefit from leveraging insights from cognitive psychology to inform medical education and
497 practice⁸⁹. How clinicians evaluate their decisions contributes to their effectiveness: An
498 overconfident clinician may overlook diagnostic possibilities, delay treatment or ignore
499 crucial information. Conversely, an underconfident clinician may be less likely to speak up in
500 a group about potential errors⁸².

501

502 Our review finds that confidence and accuracy are rarely aligned during diagnoses. Notably,
503 miscalibration of confidence is not only a function of social and environmental factors:
504 Miscalibration is consistently observed in vignette studies performed by individual medical
505 professionals, who tend to be overconfident particularly when dealing with complex cases.
506 Nevertheless, social and environmental factors may amplify systematic tendencies toward
507 misaligned confidence/certainty. Overconfidence is associated with overlooking
508 differentials, ignoring important patient information and being less willing to admit
509 mistakes. Hence, mitigating overconfidence is an important direction for future research.
510 Underconfidence has received less attention, but is observed in medical trainees^{31,33,34} and
511 can lead to negative outcomes such as delayed treatment⁹⁰ and ordering of unnecessary
512 tests⁵¹. Interventions have been tested to improve confidence calibration (such as
513 considering alternative diagnoses and guided reflection), but these have not proven fully
514 successful^{68,69}. More work is therefore needed to design interventions to improve
515 calibration. Findings from studies of “metacognition” already inform educational practices
516 outside medicine to improve student learning^{91,92}. Although cognitive interventions such as
517 considering alternative diagnoses and guided reflections have been tested, there is yet to be
518 a standardised cognitive framework to teach non-technical skills such as constructive
519 confrontation or expressions of uncertainty.

520

521 Our conceptual model of the diagnostic decision process reflects how different levels of
522 factors (related to clinician and environmental context) differently impact accuracy and
523 confidence. This model demonstrates the importance of both behavioural and work system
524 factors within healthcare and how environmental aspects can inform an individual’s
525 decision process. Considering work context is important given findings of lower confidence
526 due to environmental factors such as shift busyness and time pressures. This corresponds
527 with other findings of stress being associated with decreased confidence for intermediate
528 levels of uncertainty⁹³ and this stress could be amplified by the healthcare environment that
529 clinicians operate in. Future interventions on diagnosis can refer to this model to

530 understand the part and context of the decision process at which they are administering the
531 intervention.

532

533 Beyond these key research themes, two further points emerge from the scope and variety
534 of work identified in our review. First, our review highlights the broad relevance of
535 confidence across different medical subdisciplines, suggesting the value of targeting
536 confidence calibration within medical education as a generally applicable approach to
537 improve diagnostic decisions. Second, in terms of methodology, we find that confidence and
538 certainty have been studied in a variety of ways (e.g. using ‘assessments’ or ‘interpretations’
539 as well as diagnostic decisions), but can sometimes be seen a primary outcome measure to
540 improve. Increasing the confidence of clinicians without considering their objective accuracy
541 may exacerbate instances of overconfidence. Future work should focus instead on
542 prompting calibrated rather than increased confidence given the aforementioned impacts
543 on patient care. As objective feedback is often unattainable in medical settings, confidence
544 could be studied using methodologies like virtual reality and high-fidelity simulations that
545 better emulate real medical settings when compared to vignette methodologies, whilst also
546 having markers of objective accuracy. Methodologically, future work can also aim to
547 determine best practices around studying diagnosis, starting with a formal assessment of
548 the quality of current research.

549

550 **Implications for Future Clinical Research**

551

552 The current healthcare context presents additional challenges to clinicians, with substantial
553 increases in clinical workload in the aftermath of the COVID-19 pandemic, and workforce
554 stress and burnout at their highest in the NHS since recording began⁹⁴. Support to enhance
555 clinical decision making through improved confidence/accuracy calibration could help to
556 relieve pressure on the frontline.

557

558 *What areas of research are still underexplored within the context of medical diagnosis?*

559

560 Through our conceptual model of the diagnostic process, we identify three primary areas for
561 future work that have been relatively underexplored to date. First, most extant research has
562 studied diagnosis as a linear process in which information is presented sequentially prior to
563 a final diagnosis, at which point confidence is assessed. This is a helpful idealisation of a
564 process that in reality is complex and dynamic, with diagnosis evolving via cycling between
565 seeking information and evaluating that information in the context of currently considered
566 diagnostic possibilities. Moreover, patients’ unexpected responses to initial treatment can
567 lead to revision of an initial diagnosis and prompt further information seeking. Effects of
568 (miscalibrated) confidence are likely to be amplified within these dynamics, such as an
569 overconfident clinician paying too much attention to evidence supporting their diagnosis
570 and neglecting opposing evidence, resulting in greater overconfidence. These dynamics
571 could be studied in more naturalistic, in situ methodologies that are closer to everyday
572 medical practice. Some methodologies permit capturing diagnostic reasoning as it evolves
573 with time and the receipt of new information, such as asking clinicians to think aloud as they
574 make diagnoses⁹⁵ or using a visual representation of clinicians’ thought processes to capture
575 paths and sources of diagnoses⁶⁶. The use of high-fidelity or virtual reality simulations are
576 useful in emulating the pressure and work environment of the clinician (which may affect

577 decision making)^{96,97}, as well as providing an actual ‘patient’ to observe (unlike in textual
578 vignettes). Another approach that maintains experimental control in a more naturalistic
579 medical workplace environment is the use of standardised patients, such that actors act as
580 patients with a preset case that clinicians treat as they would regular patients^{22,98}. Use of
581 such paradigms would improve the generalisability of results.

582
583 Secondly, the vast majority of studies have investigated confidence from the perspective of
584 individual clinicians. However, diagnosis and treatment decisions are often made by teams
585 rather than individuals, particularly in secondary care settings. Evidence from organisational
586 psychology indicates that group decisions depend critically on communicated confidence
587 and uncertainty²¹: Overconfident team members can anchor a group on an incorrect
588 decision⁹⁹. Conversely, underconfident team members may fail to share important
589 information that is unknown to the rest of the group, exacerbating the problems of ‘hidden
590 information’ and ‘shared information bias’^{83,100}. In addition, clinicians may modify how they
591 communicate certainty with others, especially given the collaborative nature of healthcare
592 and the social benefits of communicating opinions with confidence in order to be listened to
593 in a group³⁴. Situational awareness (SA) is also important in a group, and higher stress may
594 be associated with overconfidence in SA¹⁰¹. Taken together, group medical decisions
595 represent an important naturalistic area for future study.

596
597 Finally, the papers in this review have also not looked at individual differences in
598 expressions of confidence, where past work from cognitive psychology has found individual
599 systematic tendencies toward higher or lower confidence^{102,103}. Hence, individual clinicians
600 may have a consistent tendency toward underconfidence or overconfidence that impacts
601 their clinical practice and that training or cognitive aids could address.

602
603

604 **CONCLUSIONS**

605
606 This scoping review indicates that, across 33 studies in which calibration of diagnostic
607 confidence to objective accuracy was measured, confidence was frequently not calibrated
608 to accuracy during diagnostic decision making processes. Notably, many factors affect
609 confidence and accuracy separately, creating instances of both overconfidence and
610 underconfidence. This miscalibration of confidence to objective accuracy can be highly
611 problematic given that confidence affects the subsequent care pathway of patients, as
612 demonstrated in studies that involved actual patients rather than patient vignettes. We
613 propose a conceptual model of diagnostic confidence that could be utilised by future
614 researchers to identify where interventions for prompting appropriate confidence can
615 target specific parts of the decision process. More holistically, the review has demonstrated
616 how diagnostic confidence is a result of the individual clinician, the patient they are treating,
617 and the environmental context the clinician works in. Future cognitive and human factors
618 interventions could target each of these levels to support more accurate decision making
619 and improved patient safety.

620
621
622 We have no competing interests to disclose. This work has not been previously published or
623 presented at a conference. This work was done was undertaken as part of a PhD at the

624 University of Oxford, which was funded by Wolfson College, Oxford via the Oxford-Wolfson
625 Marriott Graduate Scholarship (Reference: SFF2122_OGSMF-OWM_1466776).

626
627
628

629 REFERENCES

630

- 631 1. McGlynn EA, McDonald KM, Cassel CK. Measurement is essential for improving
632 diagnosis and reducing diagnostic error: a report from the Institute of Medicine.
633 [Jama](#). 2015 Dec 15;314(23):2501-2.
- 634 2. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, Elstein AS, Hasler S,
635 Kabongo ML, Krosnjar N, Odwazny R. Diagnostic error in medicine: analysis of 583
636 physician-reported errors. [Archives of internal medicine](#). 2009 Nov 9;169(20):1881-7.
- 637 3. Berwick DM, Hackbarth AD. Eliminating waste in US health care. [Jama](#). 2012 Apr
638 11;307(14):1513-6.
- 639 4. Hautz WE, Kämmer JE, Hautz SC, Sauter TC, Zwaan L, Exadaktylos AK, Birrenbach T,
640 Maier V, Müller M, Schaubert SK. Diagnostic error increases mortality and length of
641 hospital stay in patients presenting through the emergency room. [Scandinavian
642 journal of trauma, resuscitation and emergency medicine](#). 2019 Dec;27:1-2.
- 643 5. Restrepo D, Armstrong KA, Metlay JP. Annals clinical decision making: avoiding
644 cognitive errors in clinical decision making. [Annals of internal medicine](#). 2020 Jun
645 2;172(11):747-51.
- 646 6. Mamede S, van Gog T, van den Berge K, Rikers RM, van Saase JL, van Guldener C,
647 Schmidt HG. Effect of availability bias and reflective reasoning on diagnostic accuracy
648 among internal medicine residents. [Jama](#). 2010 Sep 15;304(11):1198-203.
- 649 7. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine.
650 [The American journal of medicine](#). 2008 May 1;121(5):S2-3.
- 651 8. Overall, J. E., & Williams, C. M. (1961). Models for medical diagnosis. *Behavioral
652 Science*, 6(2), 134-141.
- 653 9. Royce CS, Hayes MM, Schwartzstein RM. Teaching critical thinking: a case for
654 instruction in cognitive biases to reduce diagnostic errors and improve patient
655 safety. *Academic Medicine*. 2019 Feb 1;94(2):187-94.
- 656 10. Fleming SM, Daw ND. Self-evaluation of decision-making: A general Bayesian
657 framework for metacognitive computation. [Psychological review](#). 2017
658 Jan;124(1):91.
- 659 11. Boldt A, Yeung N. Shared neural markers of decision confidence and error detection.
660 *Journal of Neuroscience*. 2015 Feb 25;35(8):3478-84.
- 661 12. Shekhar M, Rahnev D. How do humans give confidence? A comprehensive
662 comparison of process models of perceptual metacognition. *Journal of Experimental
663 Psychology: General*. 2023 Dec 14.
- 664 13. Kiani R, Corthell L, Shadlen MN. Choice certainty is informed by both evidence and
665 decision time. *Neuron*. 2014 Dec 17;84(6):1329-42.
- 666 14. Ko YH, Feuerriegel D, Turner W, Overhoff H, Niessen E, Stahl J, Hester R, Fink GR,
667 Weiss PH, Bode S. Divergent effects of absolute evidence magnitude on decision
668 accuracy and confidence in perceptual judgements. *Cognition*. 2022 Aug
669 1;225:105125.

- 670 15. Rouault M, Seow T, Gillan CM, Fleming SM. Psychiatric symptom dimensions are
671 associated with dissociable shifts in metacognition but not task performance.
672 *Biological psychiatry*. 2018 Sep 15;84(6):443-51.
- 673 16. Schaefer PS, Williams CC, Goodie AS, Campbell WK. Overconfidence and the big five.
674 *Journal of research in Personality*. 2004 Oct 1;38(5):473-80.
- 675 17. Syzmanowicz A, Furnham A. Gender differences in self-estimates of general,
676 mathematical, spatial and verbal intelligence: Four meta analyses. *Learning and*
677 *individual Differences*. 2011 Oct 1;21(5):493-504.
- 678 18. See KE, Morrison EW, Rothman NB, Soll JB. The detrimental effects of power on
679 confidence, advice taking, and accuracy. *Organizational behavior and human*
680 *decision processes*. 2011 Nov 1;116(2):272-85.
- 681 19. Desender K, Boldt A, Yeung N. Subjective confidence predicts information seeking in
682 decision making. *Psychological science*. 2018 May;29(5):761-78.
- 683 20. Zarnoth P, Sniezek JA. The social influence of confidence in group decision making.
684 *Journal of Experimental Social Psychology*. 1997 Jul 1;33(4):345-66.
- 685 21. Silver I, Mellers BA, Tetlock PE. Wise teamwork: Collective confidence calibration
686 predicts the effectiveness of group discussion. *Journal of Experimental Social*
687 *Psychology*. 2021 Sep 1;96:104157.
- 688 22. Kovacs RJ, Lagarde M, Cairns J. Overconfident health workers provide lower quality
689 healthcare. *Journal of Economic Psychology*. 2020 Jan 1;76:102213.
- 690 23. Price PC, Stone ER. Intuitive evaluation of likelihood judgment producers: Evidence
691 for a confidence heuristic. [Journal of Behavioral Decision Making](#). 2004 Jan;17(1):39-
692 57.
- 693 24. Jaspán O, Wysocka A, Sanchez C, Schweitzer AD. Improving the relationship between
694 confidence and competence: implications for diagnostic radiology training from the
695 psychology and medical literature. [Academic Radiology](#). 2022 Mar 1;29(3):428-38.
- 696 25. Miller DJ, Spengler ES, Spengler PM. A meta-analysis of confidence and judgment
697 accuracy in clinical decision making. *Journal of Counseling Psychology*. 2015
698 Oct;62(4):553.
- 699 26. Pouget A, Drugowitsch J, Kepecs A. Confidence and certainty: distinct probabilistic
700 quantities for different goals. *Nature neuroscience*. 2016 Mar;19(3):366-74.
- 701 27. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil, H. Chapter 11:
702 Scoping Reviews (2020 version). In: Aromataris E, Munn Z (Editors). , JBI, 2020.
703 Available JBI Manual for Evidence Synthesis from <https://synthesismanual.jbi.glo>
- 704 28. Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-
705 informed management knowledge by means of systematic review. *British journal of*
706 *management*. 2003 Sep;14(3):207-22.
- 707 29. Webster J, Watson RT. Analyzing the past to prepare for the future: Writing a
708 literature review. *MIS quarterly*. 2002 Jun 1:xiii-xiii.
- 709 30. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package
710 and Shiny app for producing PRISMA 2020-compliant flow diagrams, with
711 interactivity for optimised digital transparency and Open Synthesis. *Campbell*
712 *systematic reviews*. 2022 Jun;18(2):e1230.
- 713 31. Yang H, Thompson C, Bland M. The effect of clinical experience, judgment task
714 difficulty and time pressure on nurses' confidence calibration in a high fidelity clinical
715 simulation. *BMC medical informatics and decision making*. 2012 Dec;12:1-9.

- 716 32. Gherman S, Philiastides MG. Neural representations of confidence emerge from the
717 process of decision formation during perceptual choices. *Neuroimage*. 2015 Feb
718 1;106:134-43.
- 719 33. Mann D. The Relationship between Diagnostic Accuracy and Confidence in Medical
720 Students. [ERIC](#)
- 721 34. Brezis M, Orkin-Bedolach Y, Fink D, Kiderman A. Does Physician's Training Induce
722 Overconfidence That Hampers Disclosing Errors?. [Journal of Patient Safety](#). 2019 Dec
723 1;15(4):296-8.
- 724 35. Friedman C, Gatti G, Elstein A, Franz T, Murphy G, Wolf F. Are clinicians correct when
725 they believe they are correct? Implications for medical decision support. In *MEDINFO*
726 2001 (pp. 454-458). IOS Press.
- 727 36. Fernández-Aguilar C, Martín-Martín JJ, Minué Lorenzo S, Fernández Ajuria A. Use of
728 heuristics during the clinical decision process from family care physicians in real
729 conditions. *Journal of Evaluation in Clinical Practice*. 2022 Feb;28(1):135-41.
- 730 37. Garbayo LS, Harris DM, Fiore SM, Robinson M, Kibble JD. A metacognitive confidence
731 calibration (MCC) tool to help medical students scaffold diagnostic reasoning in
732 decision-making during high-fidelity patient simulations. *Advances in Physiology*
733 *Education*. 2023 Mar 1;47(1):71-81.
- 734 38. Yang H, Thompson C. Nurses' risk assessment judgements: A confidence calibration
735 study. *Journal of Advanced Nursing*. 2010 Dec;66(12):2751-60.
- 736 39. Clayton DA, Eguchi MM, Kerr KF, Miyoshi K, Brunyé TT, Drew T, Weaver DL, Elmore
737 JG. Are Pathologists Self-Aware of Their Diagnostic Accuracy? *Metacognition and the*
738 *Diagnostic Process in Pathology*. *Medical Decision Making*. 2023 Feb;43(2):164-74.
- 739 40. Brannon LA, Carson KL. Nursing expertise and information structure influence
740 medical decision making. [Applied Nursing Research](#). 2003 Nov 1;16(4):287-90.
- 741 41. Tabak N, Bar-Tal Y, Cohen-Mansfield J. Clinical decision making of experienced and
742 novice nurses. [Western Journal of Nursing Research](#). 1996 Oct;18(5):534-47.
- 743 42. Kostopoulou O, Russo JE, Keenan G, Delaney BC, Douiri A. Information distortion in
744 physicians' diagnostic judgments. *Medical Decision Making*. 2012 Nov;32(6):831-9.
- 745 43. Hautz WE, Schubert S, Schaubert SK, Kunina-Habenicht O, Hautz SC, Kämmer JE, Eva
746 KW. Accuracy of self-monitoring: does experience, ability or case difficulty matter?.
747 [Medical education](#). 2019 Jul;53(7):735-44.
- 748 44. Mamede S, Zandbergen A, de Carvalho-Filho MA, Choi G, Goeijenbier M, van Ginkel
749 J, Zwaan L, Paas F, Schmidt HG. Role of knowledge and reasoning processes as
750 predictors of resident physicians' susceptibility to anchoring bias in diagnostic
751 reasoning: a randomised controlled experiment. *BMJ Quality & Safety*. 2024 Feb 16.
- 752 45. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy,
753 confidence, and resource requests: a vignette study. *JAMA internal medicine*. 2013
754 Nov 25;173(21):1952-8.
- 755 46. Hausmann D, Kiesel V, Zimmerli L, Schlatter N, von Gunten A, Watteringer N,
756 Rosemann T. Sensitivity for multimorbidity: The role of diagnostic uncertainty of
757 physicians when evaluating multimorbid video case-based vignettes. *PLoS one*. 2019
758 Apr 10;14(4):e0215049.
- 759 47. Li S, Zheng J, Lajoie SP. The relationship between cognitive engagement and
760 students' performance in a simulation-based training environment: an information-
761 processing perspective. *Interactive Learning Environments*. 2023 Apr 3;31(3):1532-
762 45.

- 763 48. Kuhn J, van den Berg P, Mamede S, Zwaan L, Bindels P, van Gog T. Improving medical
764 residents' self-assessment of their diagnostic accuracy: does feedback help?.
765 [Advances in Health Sciences Education](#). 2022 Mar;27(1):189-200.
- 766 49. Staal J, Katarya K, Speelman M, Brand R, Alsmas J, Sloane J, Van den Broek WW,
767 Zwaan L. Impact of performance and information feedback on medical interns'
768 confidence–accuracy calibration. [Advances in Health Sciences Education](#). 2023 Jun
769 17:1-7.
- 770 50. Soares III WE, Price LL, Prast B, Tarbox E, Mader TJ, Blanchard R. Accuracy screening
771 for ST elevation myocardial infarction in a task-switching simulation. *Western*
772 *Journal of Emergency Medicine*. 2019 Jan;20(1):177.
- 773 51. Gupta AB, Greene MT, Fowler KE, Chopra VI. Associations Between Hospitalist Shift
774 Busyness, Diagnostic Confidence, and Resource Utilization: A Pilot Study. *Journal of*
775 *Patient Safety*. 2023 Oct 1;19(7):447-52.
- 776 52. Bergl PA, Shukla N, Shah J, Khan M, Patel JJ, Nanchal RS. Factors influencing
777 diagnostic accuracy among intensive care unit clinicians—an observational study.
778 *Diagnosis*. 2024 Feb 19;11(1):31-9.
- 779 53. Mackenzie R, Dixon AK, Keene GS, Hollingworth W, Lomas DJ, Villar RN. Magnetic
780 resonance imaging of the knee: assessment of effectiveness. *Clinical radiology*. 1996
781 Apr 1;51(4):245-50.
- 782 54. Albrechtsen SS, Riis RG, Amiri M, Tanum G, Bergdal O, Blaabjerg M, Simonsen CZ,
783 Kondziella D. Impact of MRI on decision-making in ICU patients with disorders of
784 consciousness. *Behavioural Brain Research*. 2022 Mar 12;421:113729.
- 785 55. Abujudeh HH, Kaewlai R, McMahan PM, Binder W, Novelline RA, Gazelle GS, Thrall
786 JH. Abdominopelvic CT increases diagnostic certainty and guides management
787 decisions: a prospective investigation of 584 patients in a large academic medical
788 center. *American Journal of Roentgenology*. 2011 Feb;196(2):238-43.
- 789 56. Harvey CJ, Halligan S, Bartram CI, Hollings N, Sahdev A, Kingston K. Evacuation
790 proctography: a prospective study of diagnostic and therapeutic effects. *Radiology*.
791 1999 Apr;211(1):223-7.
- 792 57. Sanger PC, Simianu VV, Gaskill CE, Armstrong CA, Hartzler AL, Lordon RJ, Lober WB,
793 Evans HL. Diagnosing surgical site infection using wound photography: a scenario-
794 based study. *Journal of the American College of Surgeons*. 2017 Jan 1;224(1):8-15.
- 795 58. Hillson SD, Connelly DP, Liu Y. The effects of computer-assisted electrocardiographic
796 interpretation on physicians' diagnostic decisions. *Medical Decision Making*. 1995
797 Jun;15(2):107-12.
- 798 59. Berner ES, Maisiak RS. Influence of case and physician characteristics on perceptions
799 of decision support systems. *Journal of the American Medical Informatics*
800 *Association*. 1999 Sep 1;6(5):428-34.
- 801 60. Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of
802 decision support systems on physician opinion. *Artificial intelligence in medicine*.
803 2005 Jan 1;33(1):25-30.
- 804 61. Neugebauer M, Ebert M, Vogelmann R. A clinical decision support system improves
805 antibiotic therapy for upper urinary tract infection in a randomized single-blinded
806 study. *BMC Health Services Research*. 2020 Dec;20:1-0.
- 807 62. Kämmer JE, Schaubert SK, Hautz SC, Stroben F, Hautz WE. Differential diagnosis
808 checklists reduce diagnostic error differentially: a randomised experiment. *Medical*
809 *education*. 2021 Oct;55(10):1172-82.

- 810 63. Kourtidis P, Nurek M, Delaney B, Kostopoulou O. Influences of early diagnostic
811 suggestions on clinical reasoning. *Cognitive Research: Principles and Implications*.
812 2022 Dec 15;7(1):103.
- 813 64. Staal J, Speelman M, Brand R, Alisma J, Zwaan L. Does a suggested diagnosis in a
814 general practitioners' referral question impact diagnostic reasoning: an experimental
815 study. *BMC Medical Education*. 2022 Apr 8;22(1):256.
- 816 65. Eva WK. The influence of differentially processing evidence on diagnostic decision-
817 making (Doctoral dissertation).
- 818 66. Feyzi-Behnagh R, Azevedo R, Legowski E, Reitmeyer K, Tseytlin E, Crowley RS.
819 Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent
820 tutoring system. *Instructional science*. 2014 Mar;42:159-81.
- 821 67. Chartan C, Singh H, Krishnamurthy P, Sur M, Meyer A, Lutfi R, Stark J,
822 Thammasitboon S. Isolating red flags to enhance diagnosis (I-RED): an experimental
823 vignette study. *International Journal for Quality in Health Care*. 2019 Oct
824 31;31(8):G97-102.
- 825 68. Lambe KA, Hevey D, Kelly BD. Guided reflection interventions show no effect on
826 diagnostic accuracy in medical students. *Frontiers in psychology*. 2018 Nov
827 23;9:285916.
- 828 69. Costa Filho GB, Moura AS, Brandão PR, Schmidt HG, Mamede S. Effects of deliberate
829 reflection on diagnostic accuracy, confidence and diagnostic calibration in
830 dermatology. *Perspectives on Medical Education*. 2019 Aug 1;8:230-6.
- 831 70. Benvenuto-Andrade C, Dusza SW, Hay JL, Agero AL, Halpern AC, Kopf AW, Marghoob
832 AA. Level of confidence in diagnosis: clinical examination versus dermoscopy
833 examination. *Dermatologic surgery*. 2006 May 1;32(5):738-44.
- 834 71. Kuhn J, Mamede S, van den Berg P, Zwaan L, van Peet P, Bindels P, van Gog T.
835 Learning deliberate reflection in medical diagnosis: does learning-by-teaching help?.
836 *Advances in Health Sciences Education*. 2023 Mar;28(1):13-26.
- 837 72. Heller RF, Saltzstein HD, Caspe WB. Heuristics in medical and non-medical decision-
838 making. [The Quarterly Journal of Experimental Psychology Section A](#). 1992
839 Feb;44(2):211-35.
- 840 73. Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of
841 error in diagnostic decision making. [Medical Decision Making](#). 1991 Dec;11(4):233-9.
- 842 74. Ben-Assuli O, Sagi D, Leshno M, Ironi A, Ziv A. Improving diagnostic accuracy using
843 EHR in emergency departments: A simulation-based study. [Journal of biomedical
844 informatics](#). 2015 Jun 1;55:31-40.
- 845 75. Tio RA, Carvalho Filho MA, de Menezes Mota MF, Santanche A, Mamede S. The
846 Effect of Information Presentation Order on Residents' Diagnostic Accuracy of Online
847 Simulated Patients With Chest Pain. [Journal of graduate medical education](#). 2022
848 Aug 1;14(4):475-81.
- 849 76. Fawver B, Thomas JL, Drew T, Mills MK, Auffermann WF, Lohse KR, Williams AM.
850 Seeing isn't necessarily believing: Misleading contextual information influences
851 perceptual-cognitive bias in radiologists. [Journal of Experimental Psychology:
852 Applied](#). 2020 Dec;26(4):579.
- 853 77. Levin PD, Idrees S, Sprung CL, Weissman C, Weiss Y, Moses AE, Benenson S.
854 Antimicrobial use in the ICU: indications and accuracy—an observational trial.
855 [Journal of hospital medicine](#). 2012 Nov;7(9):672-8.

- 856 78. Calman NS, Hyman RB, Licht W. Variability in consultation rates and practitioner
857 level of diagnostic certainty. [J Fam Pract](#). 1992 Jul 1;35(1):31-8.
- 858 79. Krupat E, Wormwood J, Schwartzstein RM, Richards JB. Avoiding premature closure
859 and reaching diagnostic accuracy: some key predictive factors. *Medical education*.
860 2017 Nov;51(11):1127-37.
- 861 80. Hageman MG, Bossen JK, King JD, Ring D. Surgeon confidence in an outpatient
862 setting. [Hand](#). 2013 Dec;8(4):430-3.
- 863 81. Thorlacius-Ussing G, Bruun M, Gjerum L, Frederiksen KS, Rhodius-Meester HF, Van
864 Der Flier WM, Waldemar G, Hasselbalch SG. Comparing a single clinician versus a
865 multidisciplinary consensus conference approach for dementia diagnostics. *Journal*
866 *of Alzheimer's Disease*. 2021 Jan 1;83(2):741-51.
- 867 82. Hémon B, Michinov E, Guy D, Mancheron P, Scipion A. Speaking up about errors in
868 routine clinical practice: a simulation-based intervention with nursing students.
869 *Clinical Simulation in Nursing*. 2020 Aug 1;45:32-41.
- 870 83. Larson JR, Christensen C, Abbott AS, Franz TM (1996) Diagnosing groups: Charting
871 the flow of information in medical decision-making teams. *Journal of Personality and*
872 *Social Psychology*, 71(2), 315-330.
- 873 84. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary
874 care—a systematic review. *Family practice*. 2008 Dec 1;25(6):400-13.
- 875 85. Graber ML, Kissam S, Payne VL, Meyer AN, Sorensen A, Lenfestey N, Tant E,
876 Henriksen K, LaBresh K, Singh H. Cognitive interventions to reduce diagnostic error: a
877 narrative review. *BMJ quality & safety*. 2012 Jul 1;21(7):535-57.
- 878 86. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with
879 medical decisions: a systematic review. *BMC medical informatics and decision*
880 *making*. 2016 Dec;16:1-4.
- 881 87. Hall KH. Reviewing intuitive decision-making and uncertainty: the implications for
882 medical education. *Medical education*. 2002 Mar;36(3):216-24.
- 883 88. Bhise V, Rajan SS, Sittig DF, Morgan RO, Chaudhary P, Singh H. Defining and
884 measuring diagnostic uncertainty in medicine: a systematic review. *Journal of*
885 *general internal medicine*. 2018 Jan;33:103-15.
- 886 89. Wears RL. Diagnosing diagnosis. *Annals of Emergency Medicine*. 2014 Dec
887 1;64(6):586-7.
- 888 90. Thompson C, Cullum N, McCaughan D, Sheldon T, Raynor P. Nurses, information use,
889 and clinical decision making—the real world potential for evidence-based decisions
890 in nursing. *Evidence-based nursing*. 2004 Jul 1;7(3):68-72.
- 891 91. Dunlosky J. Strengthening the student toolbox: Study strategies to boost learning.
892 *American Educator*. 2013;37(3):12-21.
- 893 92. Putnam AL, Sungkhasettee VW, Roediger III HL. Optimizing learning in college: Tips
894 from cognitive psychology. *Perspectives on Psychological Science*. 2016
895 Sep;11(5):652-60.
- 896 93. Heereman J, Walla P. Stress, uncertainty and decision confidence. *Applied*
897 *psychophysiology and biofeedback*. 2011 Dec;36:273-9.
- 898 94. The state of Medical Education and practice in the UK - GMC. Available at:
899 [https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/the-state-](https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/the-state-of-medical-education-and-practice-in-the-uk)
900 [of-medical-education-and-practice-in-the-uk](https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/the-state-of-medical-education-and-practice-in-the-uk) (Accessed: 06 September 2024).
- 901 95. Coderre S, Mandin HH, Harasym PH, Fick GH. Diagnostic reasoning strategies and
902 diagnostic success. *Medical education*. 2003 Aug;37(8):695-703.

903 96. Schmidt E, Goldhaber-Fiebert SN, Ho LA, McDonald KM. Simulation exercises as a
904 patient safety strategy: a systematic review. *Annals of internal medicine*. 2013 Mar
905 5;158(5_Part_2):426-32.

906 97. Jans C, Bogossian F, Andersen P, Levett-Jones T. Examining the impact of virtual
907 reality on clinical decision making—An integrative review. *Nurse Education Today*.
908 2023 Jun 1;125:105767.

909 98. Kostopoulou O, Porat T, Corrigan D, Mahmoud S, Delaney BC. Diagnostic accuracy of
910 GPs when using an early-intervention decision support system: a high-fidelity
911 simulation. *British Journal of General Practice*. 2017 Mar 1;67(656):e201-8.

912 99. Mahmoodi A, Bang D, Olsen K, Zhao YA, Shi Z, Broberg K, Safavi S, Han S, Nili
913 Ahmadabadi M, Frith CD, Roepstorff A. Equality bias impairs collective decision-
914 making across cultures. *Proceedings of the National Academy of Sciences*. 2015 Mar
915 24;112(12):3835-40.

916 100. Stasser G, Titus W. Pooling of unshared information in group decision
917 making: Biased information sampling during discussion. *Journal of personality and
918 social psychology*. 1985 Jun;48(6):1467.

919 101. Price T, Tenan M, Head J, Maslin W, LaFiandra M. Acute stress causes over
920 confidence in situation awareness. In 2016 IEEE International Multi-Disciplinary
921 Conference on Cognitive Methods in Situation Awareness and Decision Support
922 (CogSIMA) 2016 Mar 21 (pp. 1-6). IEEE.

923 102. Ais J, Zylberberg A, Barttfeld P, Sigman M. Individual consistency in the
924 accuracy and distribution of confidence judgments. *Cognition*. 2016 Jan 1;146:377-
925 86.

926 103. Navajas J, Hindocha C, Foda H, Keramati M, Latham PE, Bahrami B. The
927 idiosyncratic nature of confidence. *Nature human behaviour*. 2017 Nov;1(11):810-8.
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949

Figure 1: Visual representation of confidence calibration when comparing objective accuracy (x-axis) to subjective confidence (y-axis). Confidence is said to be calibrated when the two are matched.

Figure 2: PRISMA Diagram of Literature Review. Created using the Shiny App from Haddaway et al. (2020).

Figure 3: Conceptual model depicting factors that impact the course of a diagnostic process, with links established between concepts based on findings from this scoping review. Factors are categorised in three levels: the level of the diagnostic decision process (bottom box, blue), the level of the clinician (middle box, pink) and the level of the environmental context within which the clinician operates (top box, yellow). Black arrows represent a progression from one concept to another. Green arrows indicate positive impacts between concepts; red arrows indicate the opposite (i.e. a negative relationship). Orange arrows represent links between concepts that we highlight for future research. Light grey boxes represent factors that are known to affect decisions and confidence within the psychology literature but are currently less understood in the context of medical decisions. Diagnostic Accuracy is represented as a link between a diagnosis and eventual patient outcomes, such that a diagnosis is inferred as being accurate or inaccurate only at the end of the decision process, based on whether a patient improves or deteriorates. We also illustrate the impact of the decision process by highlighting (using darker boxes and thicker arrows) the interaction between confidence, the diagnosis/decision and information gathering.

Figure 1: Visual representation of confidence calibration when comparing objective accuracy (x-axis) to subjective confidence (y-axis). Confidence is said to be calibrated when the two are matched.

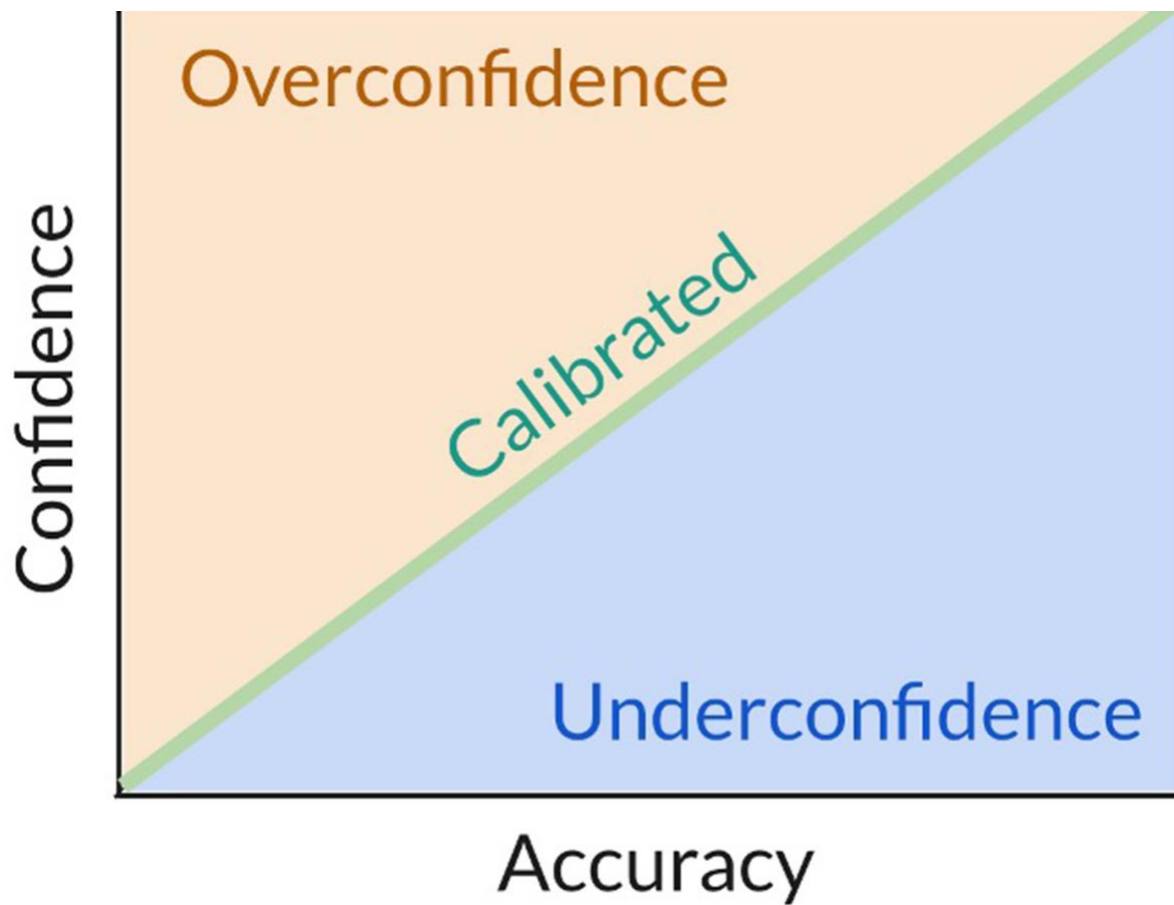


Figure 2: PRISMA Diagram of Literature Review. Created using the Shiny App from Haddaway et al. (2020).

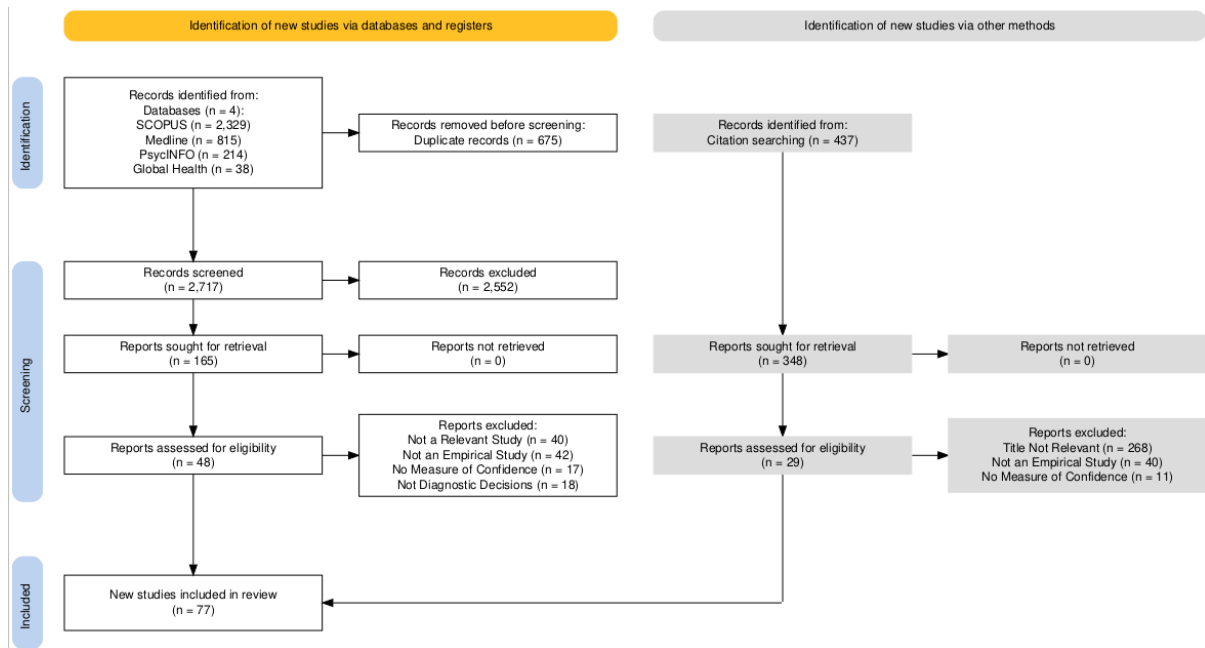


Figure 3: Conceptual model depicting factors that impact the course of a diagnostic process, with links established between concepts based on findings from this scoping review. Factors are categorised in three levels: the level of the diagnostic decision process (bottom box, blue), the level of the clinician (middle box, pink) and the level of the environmental context within which the clinician operates (top box, yellow). Black arrows represent a progression from one concept to another. Green arrows indicate positive impacts between concepts; red arrows indicate the opposite (i.e. a negative relationship). Orange arrows represent links between concepts that we highlight for future research. Light grey boxes represent factors that are known to affect decisions and confidence within the psychology literature but are currently less understood in the context of medical decisions. Diagnostic Accuracy is represented as a link between a diagnosis and eventual patient outcomes, such that a diagnosis is inferred as being accurate or inaccurate only at the end of the decision process, based on whether a patient improves or deteriorates. We also illustrate the impact of the decision process by highlighting (using darker boxes and thicker arrows) the interaction between confidence, the diagnosis/decision and information gathering.

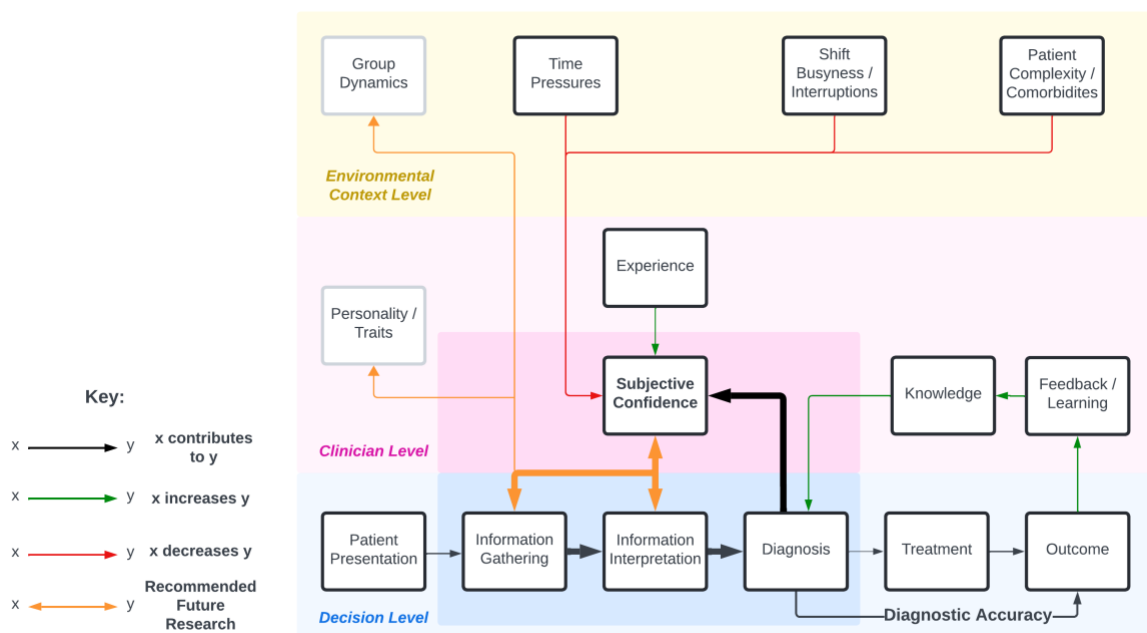


TABLE S1: Full List of Extracted Fields from Papers

Field Name	Description
Author	Listed authors on the paper
Title	Title of the published paper
Publication Title	Journal/venue that the paper was published in
Publication Year	Year that the paper was published in the above journal
Use of Vignettes	Does the paper use vignettes in their methodology?
Diagnosis Scenario/Methodology Type	Information on the type/number of vignettes/scenarios used for the study
Confidence/Calibration Measure	How confidence was measured (type of scale used for participants)
Fixed Information	Is the amount of information available the same for all study participants?
Information Stages	Is information presented to participants in a series of fixed/discrete stages? (e.g. History, Physical Examinations, Testing)
Information Seeking as a DV	Is information seeking measured in the study, either as a binary measure (i.e. asking participants whether they would seek further tests) or as a continuous measure (i.e. recording the amount of the information sought by participants)?
Novices and Experts	Does the study involve both novice/student and more experienced participants?
Experience as an IV/DV	Is the experience of the participant treated as a variable of interest, either as a categorical variable for different experience groups or as a numerical variable (e.g. number of years of experience/education)?
Multiple Differentials as DV	Are participants able to record multiple diagnoses during the study (rather than a single diagnosis)?
Case Complexity	Is the complexity/difficulty of case manipulated as a variable of interest?
Medical Discipline	Which medical discipline/subpopulation is being studied?
Any IDs	Do researchers look at individual differences between participants?
Experience Mix	Does the study use participants from a mixture of different levels of experience?
Sample Size	What was the reported sample size for the study?
Individuals vs Case	Is the study focused on investigating differences between individuals or between cases (e.g. complexity, type of patient etc)?
Accuracy Measure Operationalisation	If accuracy is measured, how is accuracy defined in this study?
Use of Decision Support Tool	Do study participants use a decision support tool?

When Confidence is Measured	At what point in the decision process is confidence elicited from participants?
How is calibration defined?	If calibration is considered (requiring both a measure of confidence and of accuracy), how is it defined as a variable?
Calibration Finding	What is the main finding from the study around the calibration of participants' diagnostic confidence?
Finding Category	Category of the study's findings (does it relate to the calibration of confidence, how confidence is used, contextual factors, the use of decision support/imaging etc.)
Finding	Summary of the study's findings from the abstract

TABLE S2: Full Table of Included Studies in Alphabetical Order (Supplemental)

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
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
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

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
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
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)

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
Berner, E.S.; Masiak, R.S.	Influence of case and physician characteristics on perceptions of decision support systems	1999	General Practice / Emergency Medicine	Written cases	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
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

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
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 electrocardiogra m interpretation	2016	GPs and Undergrads	ECG interpretation	Self-rated confidence 1-10
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

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
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 deidentified	Qualitative labels translated into %
Clayton, Dayna A.; Eguchi, Megan M.; Kerr, Kathleen F.; Miyoshi, Kiyofumi; Bruny�, 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

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 performan ce
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
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	Dermatological slides	Scale from -1 to +1

<p>Davis, D.P.; Campbell, C.J.; Poste, J.C.; Ma, G.</p>	<p>The association between operator confidence and accuracy of ultrasonography performed by novice emergency physicians</p>	<p>2005</p>	<p>Emergency Medicine</p>	<p>Ultrasound scanning</p>	<p>1-10 scale of confidence of correct test identification</p>
<p>Dreiseitl, S.; Binder, M.</p>	<p>Do physicians value decision support? A look at the effect of decision support systems on physician opinion</p>	<p>2005</p>	<p>Dermatology</p>	<p>25 dermoscopic lesions</p>	<p>1-10 scale of benign to malignant, with higher values interpreted as confident?</p>
<p>Eva, Wayne Kevin</p>	<p>The influence of differentially processing evidence on diagnostic decision-making</p>	<p>2001</p>	<p>Medical Students</p>	<p>Presenting case histories</p>	<p>Probability ratings</p>

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
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
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	Dermatological slides	6 point scale confidence in correct diagnosis

<p>Frey, J.; Braun, L. T.; Handgriff, L.; Kendziora, B.; Fischer, M. R.; Reincke, M.; Zwaan, L.; Schmidmaier, R.</p>	<p>Insights into diagnostic errors in endocrinology: a prospective, case-based, international study**</p>	<p>2023</p>	<p>Endocrinology</p>	<p>5 patient cases</p>	<p>1-10 confidence in diagnosis</p>
<p>Friedman, C.; Gatti, G.; Elstein, A.; Franz, T.; Murphy, G.; Wolf, F.</p>	<p>Are clinicians correct when they believe they are correct? Implications for medical decision support</p>	<p>2001</p>	<p>Internal Medicine</p>	<p>36 clinical cases split into 4 equal groups</p>	<p>Confidence in each diagnosis</p>

<p>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.</p>	<p>Do physicians know when their diagnoses are correct?: Implications for decision support and error reduction</p>	<p>2005</p>	<p>Internal Medicine</p>	<p>2-4 page medical synopses diagnosis</p>	<p>Likelihood to seek assistance to reach a diagnosis</p>
<p>Garbayo, Luciana S.; Harris, David M.; Fiore, Stephen M.; Robinson, Matthew; Kibble, Jonathan D.</p>	<p>A metacognitive confidence calibration (MCC) tool to help medical students scaffold diagnostic reasoning in decision-making during high-fidelity patient simulations</p>	<p>2023</p>	<p>Medical Students</p>	<p>High Fidelity Sim (Cases: Heart Failure, Respiratory Distress, DKA, heat exhaustion)</p>	<p>7 point likert scale of confidence</p>

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
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
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
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

Hausmann, D.; Kiesel, V.; Zimmerli, L.; Schlatter, N.; von Gunten, A.; Wattering, 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
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
Hautz, Wolf E; Schubert, Sebastian; Schauber, Stefan K; Kunina_Haben icht, 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%)

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
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
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 Medicine	6 clinical scenarios	10 point scale of confidence

<p>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.</p>	<p>Assessment of a Diagnostic Classification System for Management of Lesions to Exclude Melanoma**</p>	<p>2021</p>	<p>Pathology / Dermatology</p>	<p>217 Lesions prepared and stained from patients</p>	<p>1-5 confidence</p>
<p>Keene, T.; Pammer, K.; Lord, B.; Shipp, C.</p>	<p>Dispatch information affects diagnosis in paramedics: an experimental study of applied dual-process theory**</p>	<p>2022</p>	<p>Paramedics</p>	<p>Vignettes in two parts with an intuitive impression and then diagnosis, with or without secondary task distraction</p>	<p>4 point scale</p>
<p>Kostopoulou, Olga; Russo, J. Edward; Keenan, Greg; Delaney, Brendan C.; Douiri, Abdel</p>	<p>Information Distortion in Physicians' Diagnostic Judgments</p>	<p>2012</p>	<p>Primary Care</p>	<p>3 clinical scenarios each with 2 competing diagnoses</p>	<p>21 point likelihood</p>

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
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
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

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
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
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 antimicrobi als

<p>Li, S.; Zheng, J.; Lajoie, S. P.</p>	<p>The relationship between cognitive engagement and students' performance in a simulation-based training environment: an information-processing perspective**</p>	<p>2023</p>	<p>Medical Students</p>	<p>Two patient cases shown</p>	<p>0-100% scale confidence in diagnosis</p>
<p>Mackenzie, R; Dixon, A K; Keene, G S; Hollingworth, W; Lomas, D J; Villar, R N</p>	<p>Magnetic resonance imaging of the knee: assessment of effectiveness.</p>	<p>1996</p>	<p>Radiology</p>	<p>Observation of knee MRI patients</p>	<p>5 point visual analogue confidence scale</p>

<p>Mamede, S.; Zandbergen, A.; De Carvalho-Filho, M.A.; Choi, G.; Goeijenbier, M.; Van Ginkel, J.; Zwaan, L.; Paas, F.; Schmidt, H.G.</p>	<p>Role of knowledge and reasoning processes as predictors of resident physicians' susceptibility to anchoring bias in diagnostic reasoning: A randomised controlled experiment</p>	<p>2024</p>	<p>Internal Medicine</p>	<p>6 clinical vignettes (with vs without salient distracting features)</p>	<p>Confidence in diagnosis</p>
<p>Mann, Doug</p>	<p>The Relationship between Diagnostic Accuracy and Confidence in Medical Students.</p>	<p>1993</p>	<p>Medical Students / Cardiac</p>	<p>ECG slides - Classification of cardiac dysrhythmias</p>	<p>11 point scale, 0-100%</p>

<p>Marx, G.; Koens, S.; Von Dem Knesebeck, O.; Scherer, M.</p>	<p>Age and gender differences in diagnostic decision-making of early heart failure: Results of a mixed- methods interview-study using video vignettes</p>	<p>2022</p>	<p>General Practice</p>	<p>Video vignettes</p>	<p>0-100% certainty</p>
<p>Maserejian, N.N.; Lutfey, K.E.; McKinlay, J.B.</p>	<p>Do physicians attend to base rates? prevalence data and statistical discrimination in the diagnosis of coronary heart disease: Physicians and coronary heart disease</p>	<p>2009</p>	<p>Primary Care</p>	<p>Vignettes of CHD</p>	<p>0-100 scale of certainty</p>

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)
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)
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)

<p>Neugebauer, M.; Ebert, M.; Vogelmann, R.</p>	<p>A clinical decision support system improves antibiotic therapy for upper urinary tract infection in a randomized single-blinded study.</p>	<p>2020</p>	<p>Medical Doctors (Internal Medicine)</p>	<p>Fictive Paper Case</p>	<p>Confidence in Diagnosis (%)</p>
<p>Oskay, A.</p>	<p>Evaluation of thoracic computed tomography interpretation by emergency medicine residents with regards to accuracy and confidence</p>	<p>2023</p>	<p>Emergency Medicine</p>	<p>30 CT scans</p>	<p>1-10 Confidence</p>

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
Redelmeier, Donald A.; Shafir, Eldar	The Fallacy of a Single Diagnosis	2023	Primary Care	Series of vignettes to diagnosis COVID	% likelihood
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)
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 identificati on

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
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
Staal, J.; Alsmas, 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

<p>Staal, J.; Katarya, K.; Speelman, M.; Brand, R.; Alsma, J.; Sloane, J.; Van den Broek, W. W.; Zwaan, L.</p>	<p>Impact of performance and information feedback on medical interns' confidence– accuracy calibration**</p>	<p>2023</p>	<p>Medical Students</p>	<p>X-ray interpretation</p>	<p>0-10 confidence in diagnosis</p>
<p>Staal, J.; Speelman, M.; Brand, R.; Alsma, J.; Zwaan, L.</p>	<p>Does a suggested diagnosis in a general practitioners' referral question impact diagnostic reasoning: an experimental study</p>	<p>2022</p>	<p>Internal Medicine</p>	<p>6 cases formatted as GP referral letters</p>	<p>0-10 confidence in diagnosis</p>
<p>Tabak, Nili; Bar-Tal, Yoram; Cohen- Mansfield, Jiska</p>	<p>Clinical decision making of experienced and novice nurses</p>	<p>1996</p>	<p>Nursing</p>	<p>Two scenarios</p>	<p>0-100% scale confidence in diagnosis</p>

<p>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.</p>	<p>Comparing a Single Clinician Versus a Multidisciplinary Consensus Conference Approach for Dementia Diagnostics**</p>	<p>2021</p>	<p>Neurology</p>	<p>Real patient evaluations</p>	<p>0-100 Visual analogue scale</p>
<p>Tio, R. A.; Filho, M. A. C.; de Menezes Mota, M. F.; Santanchè, A.; Mamede, S.</p>	<p>The Effect of Information Presentation Order on Residents' Diagnostic Accuracy of Online Simulated Patients With Chest Pain**</p>	<p>2022</p>	<p>Cardiology</p>	<p>12 clinical cases presented in 2 diagnostic rounds (history and EKG)</p>	<p>0-100 confidence</p>
<p>van Hout, H.P.J.; Vernooij-Dassen, M.J.; Stalman, W.A.B.</p>	<p>Diagnosing dementia with confidence by GPs</p>	<p>2007</p>	<p>General Practice</p>	<p>Observation of dementia patients</p>	<p>4 point likert scale</p>

<p>van Sassen, C.; Mamede, S.; Bos, M.; van den Broek, W.; Bindels, P.; Zwaan, L.</p>	<p>Do malpractice claim clinical case vignettes enhance diagnostic accuracy and acceptance in clinical reasoning education during GP training? **</p>	<p>2023</p>	<p>General Practice</p>	<p>Cases with and without malpractice claim information</p>	<p>0-100 confidence</p>
<p>Wood, Greg; Batt, Jeremy; Appelboam, Andrew; Harris, Adrian; Wilson, Mark R.</p>	<p>Exploring the Impact of Expertise, Clinical History, and Visual Search on Electrocardiogra m Interpretation **</p>	<p>2013</p>	<p>ED</p>	<p>ECG traces and eye tracking</p>	<p>1-10 confidence in diagnosis</p>

<p>Yang, H.; Thompson, C.; Bland, M.</p>	<p>Effect of improving the realism of simulated clinical judgement tasks on nurses' overconfidence and underconfidence: Evidence from a comparative confidence calibration analysis**</p>	<p>2012</p>	<p>Nursing</p>	<p>Both paper and high fidelity sim scenarios</p>	<p>0-100 confidence</p>
<p>Yang, Huiqin; Thompson, Carl</p>	<p>Nurses' risk assessment judgements: a confidence calibration study: Nurses' risk assessment judgements</p>	<p>2010</p>	<p>Nursing</p>	<p>Risk assessment vignettes</p>	<p>0-100 confidence</p>

<p>Yang, Huiqin; Thompson, Carl; Bland, Martin</p>	<p>The effect of clinical experience, judgment task difficulty and time pressure on nurses' confidence calibration in a high fidelity clinical simulation</p>	<p>2012</p>	<p>Nursing</p>	<p>High Fidelity Sim</p>	<p>0-100 confidence</p>
--	---	-------------	----------------	--------------------------	-----------------------------

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