Scoping Review - Full Table of Included Papers

Papers are arranged in alphabetical order. Studies marked with ** next to their title were included via citation tracking.

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

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
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.; Maisiak, R.S.	Influence of case and physician characteristics on perceptions of decision support systems	1999	General Practice / Emergency Medicine	Written cases	1-5 confidence
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

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Cairns, A.W.; Bond, R.R.; Finlay, D.D.; Breen, C.; Guldenring, D.; Gaffney, R.; Gallagher, A.G.; Peace, A.J.; Henn, P.	A computer-human interaction model to improve the diagnostic accuracy and clinical decision-making during 12-lead electrocardiogram interpretation	2016	GPs and Undergrads	ECG interpretation	Self-rated confidence 1-10
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 deindentified	Qualitative labels translated into %

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Clayton, Dayna A.; Eguchi, Megan M.; Kerr, Kathleen F.; Miyoshi, Kiyofumi; Brunyé, Tad T.; Drew, Trafton; Weaver, Donald L.; Elmore, Joann G.	Are Pathologists Self-Aware of Their Diagnostic Accuracy? Metacognition and the Diagnostic Process in Pathology	2023	Pathology	Diagnosis based on slides for microscopes	6 point scale confidence in correct diagnosis
Cleary, T. J.; Konopasky, A.; La Rochelle, J. S.; Neubauer, B. E.; Durning, S. J.; Artino, A. R.	First-year medical students' calibration bias and accuracy across clinical reasoning activities**	2019	Medical Students	Some of kind of virtual patient sim	Estimations of performance
Costa Filho, G. B.; Moura, A. S.; Brandão, P. R.; Schmidt, H. G.; Mamede, S.	Effects of deliberate reflection on diagnostic accuracy, confidence and diagnostic calibration in dermatology**	2019	Medical Students / dermatology	12 dermatological images	0-100% scale confidence in diagnosis
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
Davis, D.P.; Campbell, C.J.; Poste, J.C.; Ma, G.	The association between operator confidence and accuracy of ultrasonography performed by novice emergency physicians	2005	Emergency Medicine	Ultrasound scanning	1-10 scale of confidence of correct test identification

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Dreiseitl, S.; Binder, M.	Do physicians value decision support? A look at the effect of decision support systems on physician opinion	2005	Dermatology	25 dermoscopic lesions	1-10 scale of benign to malignant, with higher values interpreted as confident?
Eva, Wayne Kevin	The influence of differentially processing evidence on diagnostic decision-making	2001	Medical Students	Presenting case histories	Probability ratings
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

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Frey, J.; Braun, L. T.; Handgriff, L.; Kendziora, B.; Fischer, M. R.; Reincke, M.; Zwaan, L.; Schmidmaier, R.	Insights into diagnostic errors in endocrinology: a prospective, case-based, international study**	2023	Endocrinology	5 patient cases	1-10 confidence in diagnosis
Friedman, C.; Gatti, G.; Elstein, A.; Franz, T.; Murphy, G.; Wolf, F.	Are clinicians correct when they believe they are correct? Implications for medical decision support	2001	Internal Medicine	36 clinical cases split into 4 equal groups	Confidence in each diagnosis
Friedman, Charles P.; Gatti, Guido G.; Franz, Timothy M.; Murphy, Gwendolyn C.; Wolf, Fredric M.; Heckerling, Paul S.; Fine, Paul L.; Miller, Thomas M.; Elstein, Arthur S.	Do physicians know when their diagnoses are correct?: Implications for decision support and error reduction	2005	Internal Medicine	2-4 page medical synopses diagnosis	Likelihood to seek assistance to reach a diagnosis
Garbayo, Luciana S.; Harris, David M.; Fiore, Stephen M.; Robinson, Matthew; Kibble, Jonathan D.	A metacognitive confidence calibration (MCC) tool to help medical students scaffold diagnostic reasoning in decision-making during high-fidelity patient simulations	2023	Medical Students	High Fidelity Sim (Cases: Heart Failure, Respiratory Distress, DKA, heat exhaustion)	7 point likert scale of confidence

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Gruppen, L; Wolf, F; Billi, J	Information Gathering and Integration as Sources of Error in Diagnostic Decision Making**	1991	Primary Care	Vignettes deciding between two diagnostic alternatives	Probability correct
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.; Wattinger, N.; Rosemann, T.	Sensitivity for multimorbidity: The role of diagnostic uncertainty of physicians when evaluating multimorbid video case-based vignettes	2019	General Practice / Emergency Medicine	Video vignettes	0-100% scale confidence in diagnosis
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

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Hautz, Wolf E; Schubert, Sebastian; Schauber, Stefan K; Kunina_Habenicht, Olga; Hautz, Stefanie C; Kämmer, Juliane E; Eva, Kevin W	Accuracy of self-monitoring: does experience, ability or case difficulty matter?	2019	Medical Students	6 clinical scenarios	10 point scale (0% to 100%)
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
Katz, I.; O'Brien, B.; Clark, S.; Thompson, C. T.; Schapiro, B.; Azzi, A.; Lilleyman, A.; Boyle, T.; Espartero, L. J. L.; Yamada, M.; Prow, T. W.	Assessment of a Diagnostic Classification System for Management of Lesions to Exclude Melanoma**	2021	Pathology / Dermatology	217 Lesions prepared and stained from patients	1-5 confidence

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Keene, T.; Pammer, K.; Lord, B.; Shipp, C.	Dispatch information affects diagnosis in paramedics: an experimental study of applied dual-process theory**	2022	Paramedics	Vignettes in two parts with an intuitive impression and then diagnosis, with or without secondary task distraction	4 point scale
Kostopoulou, Olga; Russo, J. Edward; Keenan, Greg; Delaney, Brendan C.; Douiri, Abdel	Information Distortion in Physicians' Diagnostic Judgments	2012	Primary Care	3 clinical scenarios each with 2 competing diagnoses	21 point likelihood
Kourtidis, Ploutarchos; Nurek, Martine; Delaney, Brendan; Kostopoulou, Olga	Influences of early diagnostic suggestions on clinical reasoning	2022	Family Medicine	2 patient scenarios with or without diagnostic suggestions	10 point visual analogue scale of certainty
Krupat, Edward; Wormwood, Jolie; Schwartzstein, Richard M; Richards, Jeremy B	Avoiding premature closure and reaching diagnostic accuracy: some key predictive factors	2017	Internal Medicine	4 complex vignettes	1-100 scale of certainty
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

Author(s)	Title	Year	Discipline	Methodology	Measure of 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
Leblanc, Vicki R.; Norman, Geoffrey R.; Brooks, Lee R.	Effect of a Diagnostic Suggestion on Diagnostic Accuracy and Identification of Clinical Features:	2001	Medical Students	Scenarios with photographs with clinical features	
Levin, P. D.; Idrees, S.; Sprung, C. L.; Weissman, C.; Weiss, Y.; Moses, A. E.; Benenson, S.	Antimicrobial use in the ICU: Indications and accuracy - an observational trial.	2012	ICU	Observational in ICU	Certainty of presence of infection when starting patients on antimicrobials

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Li, S.; Zheng, J.; Lajoie, S. P.	The relationship between cognitive engagement and students' performance in a simulation-based training environment: an information-processing perspective**	2020	Medical Students	Two patient cases shown	0-100% scale confidence in diagnosis
Mackenzie, R; Dixon, A K; Keene, G S; Hollingworth, W; Lomas, D J; Villar, R N	Magnetic resonance imaging of the knee: assessment of effectiveness.	1996	Radiology	Observation of knee MRI patients	5 point visual analogue confidence scale
Mamede, S.; Zandbergen, A.; De Carvalho-Filho, M.A.; Choi, G.; Goeijenbier, M.; Van Ginkel, J.; Zwaan, L.; Paas, F.; Schmidt, H.G.	Role of knowledge and reasoning processes as predictors of resident physicians' susceptibility to anchoring bias in diagnostic reasoning: A randomised controlled experiment	2024	Internal Medicine	6 clinical vignettes (with vs without salient distracting features)	Confidence in diagnosis
Mann, Doug	The Relationship between Diagnostic Accuracy and Confidence in Medical Students.	1993	Medical Students / Cardiac	ECG slides - Classification of cardiac dysrhythmias	11 point scale, 0-100%

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

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Neuge bauer, M.; Ebert, M.; Vogelmann, R.	A clinical decision support system improves antibiotic therapy for upper urinary tract infection in a randomized single-blinded study.	2020	Medical Doctors (Internal Medicine)	Fictive Paper Case	Confidence in Diagnosis (%)
Oskay, A.	Evaluation of thoracic computed tomography interpretation by emergency medicine residents with regards to accuracy and confidence	2023	Emergency Medicine	30 CT scans	1-10 Confidence
Pusic, M. V.; Chiaramonte, R.; Gladding, S.; Andrews, J. S.; Pecaric, M. R.; Boutis, K.	Accuracy of self-monitoring during learning of radiograph interpreta-tion**	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	% likelhiood
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)

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Schoenherr, Jordan Richard; Waechter, Jason; Millington, Scott J	Subjective awareness of ultrasound expertise development: individual experience as a determinant of overconfidence	2018	Cardiology	Cardiac ultrasound case studies	6 point scale confidence in correct identification
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.; Alsma, J.; Mamede, S.; Olson, A. P. J.; Prins-van Gilst, G.; Geerlings, S. E.; Plesac, M.; Sundberg, M. A.; Frens, M. A.; Schmidt, H. G.; Van den Broek, W. W.; Zwaan, L.	The relationship between time to diagnose and diagnostic accuracy among internal medicine residents: a randomized experiment**	2021	Internal Medicine	8 clinical case	0-100% scale confidence that diagnosis was correct
Staal, J.; Katarya, K.; Speelman, M.; Brand, R.; Alsma, J.; Sloane, J.; Van den Broek, W. W.; Zwaan, L.	Impact of performance and information feedback on medical interns' confidence–accuracy calibration**	2023	Medical Students	X-ray interpretation	0-10 confidence in diagnosis

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Staal, J.; Speelman, M.; Brand, R.; Alsma, J.; Zwaan, L.	Does a suggested diagnosis in a general practitioners' referral question impact diagnostic reasoning: an experimental study	2022	Internal Medicine	6 cases formatted as GP referral letters	0-10 confidence in diagnosis
Tabak, Nili; Bar-Tal, Yoram; Cohen-Mansfield, Jiska	Clinical decision making of experienced and novice nurses	1996	Nursing	Two scenarios	0-100% scale confidence in diagnosis
Thorlacius-Ussing, G.; Bruun, M.; Gjerum, L.; Frederiksen, K. S.; Rhodius-Meester, H. F. M.; Van Der Flier, W. M.; Waldemar, G.; Hasselbalch, S. G.; Nobili, F.	Comparing a Single Clinician Versus a Multi-disciplinary Consensus Conference Approach for Dementia Diagnostics**	2021	Neurology	Real patient evaluations	0-100 Visual analogue scale
Tio, R. A.; Filho, M. A. C.; de Menezes Mota, M. F.; Santanchè, A.; Mamede, S.	The Effect of Information Presentation Order on Residents' Diagnostic Accuracy of Online Simulated Patients With Chest Pain**	2022	Cardiology	12 clinical cases presented in 2 diagnostic rounds (history and EKG)	0-100 confidence

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Trueblood, Jennifer S.; Eichbaum, Quentin; Seegmiller, Adam C.; Stratton, Charles; O'Daniels, Payton; Holmes, William R.	Disentangling prevalence induced biases in medical image decision-making	2021	Medical Students / Imaging	Cell scans (cancer identification)	
van Hout, H.P.J.; Vernooij-Dassen, M.J.; Stalman, W.A.B.	Diagnosing dementia with confidence by GPs	2007	General Practice	Observation of dementia patients	4 point likert scale
van Sassen, C.; Mamede, S.; Bos, M.; van den Broek, W.; Bindels, P.; Zwaan, L.	Do malpractice claim clinical case vignettes enhance diagnostic accuracy and acceptance in clinical reasoning education during GP training?**	2023	General Practice	Cases with and without malpractice claim information	0-100 confidence
Wood, Greg; Batt, Jeremy; Appelboam, Andrew; Harris, Adrian; Wilson, Mark R.	Exploring the Impact of Expertise, Clinical History, and Visual Search on Electrocardiogram Interpretation**	2014	ED	ECG traces and eye tracking	1-10 confidence in diagnosis

Author(s)	Title	Year	Discipline	Methodology	Measure of Confidence
Yang, H.; Thompson, C.; Bland, M.	Effect of improving the realism of simulated clinical judgement tasks on nurses' overconfidence and underconfidence: Evidence from a comparative confidence calibration analysis**	2012	Nursing	Both paper and high fidelity sim scenarios	0-100 confidence
Yang, Huiqin; Thompson, Carl	Nurses' risk assessment judgements: a confidence calibration study: Nurses' risk assessment judgements	2010	Nursing	Risk assessment vignettes	0-100 confidence
Yang, Huiqin; Thompson, Carl; Bland, Martin	The effect of clinical experience, judgment task difficulty and time pressure on nurses' confidence calibration in a high fidelity clinical simulation	2012	Nursing	High Fidelity Sim	0-100 confidence

Vignette Marking Scheme (Studies 2 and 3)

Condition	Abbreviation	Presenting Complaint	Accepted Answers
Temporal Arteritis	TA	Patient is a 68 year old male presented with fever and arthralgia.	Any inflammatory arthritis is accepted
Ulcerative Colitis	UC	Patient is a 60 year old male presented with 2 day history of bloody diarrhoea.	Infectious colitis, ischemic colitis and diverticulitis are also accepted answers.

Condition	Abbreviation	Presenting Complaint	Accepted Answers
Miliary Tuberculosis	MTB	Patient is a 62 year old male admitted for fevers and generalised weakness.	Any TB or lymphoma type is accepted
Aortic Dissection	AD	Patient is a 58 year old female presented with shortness of breath.	Pulmonary embolism and coarctation of the aorta are also accepted answers. Aortic stenosis
Guillain-Barré Syndrome	GBS	Patient is a 67 year old male presented with weakness of the legs for 24 hours.	Cauda Equina Syndrome is also accepted
Thrombotic Thrombocytopenic Purpura	TTP	Patient is a 20 year old male was admitted from an outside hospital with complaints of a headache and slurred speech.	ITP or Meningitis are also accepted.

Table S1: Marking scheme used to denote differentials that are considered as correct for each of the six patient cases/vignettes. The same marking scheme is applied for online and think-aloud vignette studies. The presenting complaint is shown to participants at the start of the case, before they start seeking information.

Vignette Information Requests

Patient History	Physical Examinations	Testing
History of Presenting Complaint	Take Pulse	Urine Dipstick
Past Medical History	Measure Blood Pressure	ECG
Medications	Assess Respiratory Rate	Abdominal CT Scan
Allergies	Auscultate Lungs	Venous Blood Gas
Family History	Auscultate the Heart	CRP and ESR
Social History	Assess Eyes	Clotting Test
	Measure Temperature	FBC
	Abdomen Examination	Other Biochemistry tests
	Rectal Examination	UREA and Electrolytes
	Neck/Throat Examination	Chest X-Ray
	Assess Head	
	Neurological Exam Record	
	Assess Extremities	

Table S2: Full list of possible information requests that participants can make. This set of information is the same for all cases. The same vignettes and corresponding information are used for the online and think-aloud vignette studies.

Calibration of Confidence to Alternative Accuracy Measures

Differential Accuracy

When comparing Differential Accuracy (if a correct differential is provided in the participant's list) to Confidence, we find, across stages, participants' Confidence was not aligned to their Accuracy. Instead, we find evidence of underconfidence at all stages. There was evidence of a significant difference between the two at the Patient History ($t(84) = 8.24$, MDiff = 0.24, $p < .001$), Physical Examination stage ($t(84) = -9.09$, MDiff = -0.25, $p < .001$), and Testing stage ($t(84) = -7.74$, MDiff = -0.22, $p < .001$).

In order to examine the observed underconfidence in more detail, we compare confidence and Differential Accuracy by case (the mean values of which can be found in Table 1 of the main thesis). We conducted paired t-tests for each condition's cases by comparing Differential Accuracy and confidence values (at the final Testing stage) to observe if they significantly differ from each other. A p value of less than .05 is interpreted as evidence for overconfidence or underconfidence (depending on the direction of the effect). We observed underconfidence for the GBS case ($t(84) = -7.43$, MDiff = -0.39, $p = < .001$), the TA case ($t(84) = -5.07$, MDiff = -0.25, $p = < .001$), the TTP case ($t(84) = -3.23$, MDiff = -0.2, $p = < .001$) and the UC case ($t(82) = -14.83$, MDiff = -0.38, $p = < .001$). The remaining cases did not yield a significant effect.

Highest Likelihood Accuracy

When comparing Highest Likelihood Accuracy (likelihood assigned to the highest likelihood differential if it is correct) to Confidence, we find, across stages, participants' Confidence was not aligned to their Accuracy. Instead, we find evidence of overconfidence at all stages. There was evidence of a significant difference between the two at the Patient History ($t(84) = -2.49$, MDiff = -0.05, $p = 0.01$), Physical Examination stages ($t(84) = 4.45$, MDiff = 0.09, $p < .001$), and Testing stage ($t(84) = 6.84$, MDiff = 0.16, $p < .001$).

In order to examine the observed overconfidence in more detail, we compare confidence and Highest Likelihood Accuracy by case (the mean values of which can be found in Table 1 of the main thesis). We conducted paired t-tests for each condition's cases by comparing Highest Likelihood Accuracy and confidence values (at the final Testing stage) to observe if they significantly differ from each other. A p value of less than .05 is interpreted as evidence for overconfidence or underconfidence (depending on the direction of the effect). We observed overconfidence for the AD case ($t(84) = 8.92$, MDiff = 0.37, $p = < .001$), the MTB case ($t(83) = 7.66$, MDiff = 0.35, $p = < .001$) and the TTP case ($t(84) = 4.09$, MDiff = 0.21, $p = < .001$). The remaining cases did not yield a significant effect.

Debrief Questionnaire from Think-Aloud Study

Each question has a corresponding follow-up question below in case they are not answered by responses to the main questions.

- 1. What's your general approach to making diagnoses? *Follow-Up:* Do you have those cognitive aids or frameworks you use?
- 2. Do you tend to keep a broad set of differentials in mind? *Follow-Up:* Are there particular situations where having a narrower set would be more useful?
- 3. How do you decide what information or tests to get on a patient? *Follow-Up:* Would you say you tend to seek information to confirm or to rule out differentials that you have in mind?
- 4. How similar was your diagnostic reasoning on this task versus how you would approach diagnosis in real life? *Follow-Up:* Was there anything that prevented you from approaching the task as you would in real life?

Diagnostic Appropriateness Marking Scheme for VR Study

Scenario	Probable/Possible Differentials	Improbable/Unlikely Differentials
Asthma	Asthma / asthma excerbation	Epiglotitis
	Pneumonia / LRTI	Croup
	RSV / Viral URTI	PE
	Foreign Body	
	Anaphylaxis	
	Viral Induced Wheeze	
DKA	DKA	Alcohol ingestion
	URTI / throat infection / tonsillitis	Sickle Cell
	Gastroenteritis / abdominal infection	Inborn errors of metabolism
	Insulin non compliance	
	Sepsis	
	Viral infection	
Seizure	Epilepsy / Febrile Seizure	Fictitious / malingering
	Meningitis / CNS infection / encephalitis	Alcohol withdrawing
	Hypo / hypoglycaemia	Sickle cell
	Non accidential injury (NEA)	Inborn errors of metabolism
	Space occupying lesion (SOL) / tumour	
Pneumonia	Pneumonia / LRTI	Anaphylaxis
	URTI / cold / flu	Pleural effusion
	Viral LRTI	Pneumothorax
	Asthma	
	Inhaled foreign body (FB)	

Table S3: Marking criteria for the VR Study. Differentials are shown for each scenario that were marked as either probable/possible and those categorised as improbable/unlikely. Any differentials not included in this table were marked as incorrect.

R Environment and Packages

```
# print("R version:")
# version$version.string
#
# print("Rstudio version:")
# rstudioversion <- rstudioapi::versionInfo()
# rstudioversion$version
#
# print("Citations for packages used:")
# get_pkgs_info(pkgs = required_packages, out.dir = getwd())
# pkgs <- scan_packages()
# get_citations(pkgs$pkg, out.dir = getwd(), include.RStudio = TRUE)
# cite_packages(pkgs = required_packages, output = "table", out.format = "Rmd", out.dir = getwd())
#
# required_packages %>%
#   map(citation) %>%
#   print(style = "text")
```