Original Investigation

Physicians' Diagnostic Accuracy, Confidence, and Resource Requests
A Vignette Study

Ashley N. D. Meyer, PhD; Velma L. Payne, PhD, MBA; Derek W. Meeks, MD; Radha Rao, MD; Hardeep Singh, MD, MPH

IMPORTANCE Little is known about the relationship between physicians' diagnostic accuracy and their confidence in that accuracy.

OBJECTIVE To evaluate how physicians' *diagnostic calibration*, defined as the relationship between diagnostic accuracy and confidence in that accuracy, changes with evolution of the diagnostic process and with increasing diagnostic difficulty of clinical case vignettes.

DESIGN, SETTING, AND PARTICIPANTS We recruited general internists from an online physician community and asked them to diagnose 4 previously validated case vignettes of variable difficulty (2 *easier*; 2 *more difficult*). Cases were presented in a web-based format and divided into 4 sequential phases simulating diagnosis evolution: history, physical examination, general diagnostic testing data, and definitive diagnostic testing. After each phase, physicians recorded 1 to 3 differential diagnoses and corresponding judgments of confidence. Before being presented with definitive diagnostic data, physicians were asked to identify additional resources they would require to diagnose each case (ie, additional tests, second opinions, curbside consultations, referrals, and reference materials).

MAIN OUTCOMES AND MEASURES Diagnostic accuracy (scored as 0 or 1), confidence in diagnostic accuracy (on a scale of 0-10), diagnostic calibration, and whether additional resources were requested (no or yes).

RESULTS A total of 118 physicians with broad geographical representation within the United States correctly diagnosed 55.3% of easier and 5.8% of more difficult cases (*P* < .001). Despite a large difference in diagnostic accuracy between easier and more difficult cases, the difference in confidence was relatively small (7.2 vs 6.4 out of 10, for easier and more difficult cases, respectively) (*P* < .001) and likely clinically insignificant. Overall, diagnostic calibration was worse for more difficult cases (*P* < .001) and characterized by overconfidence in accuracy. Higher confidence was related to decreased requests for additional diagnostic tests (*P* = .01); higher case difficulty was related to more requests for additional reference materials (*P* = .01).

CONCLUSIONS AND RELEVANCE Our study suggests that physicians' level of confidence may be relatively insensitive to both diagnostic accuracy and case difficulty. This mismatch might prevent physicians from reexamining difficult cases where their diagnosis may be incorrect.

← Invited Commentary
page 1959

+ Supplemental content at
jamainternalmedicine.com

Author Affiliations: Houston Veterans Affairs Health Services Research and Development Center of Excellence and the Section of Health Services Research, Houston, Texas (Meyer, Payne, Meeks, Singh); Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas (Meyer, Payne, Meeks, Rao, Singh); Baylor College of Medicine, Houston, Texas (Meyer, Payne, Meeks, Rao, Singh).

Corresponding Author: Hardeep Singh, MD, MPH, Veterans Affairs Medical Center (152), 2002 Holcombe Blvd, Houston, TX 77030 (hardeeps@bcm.edu).

Diagnostic errors can lead to patient harm[1] but have received inadequate exploration relative to other patient safety concerns.[2-4] Physician overconfidence is thought to be one of many contributing factors to diagnostic error and occurs when the relationship between accuracy and confidence is miscalibrated or misaligned such that confidence is higher than it should be.[5] The relationship between diagnostic accuracy and confidence is unclear,[6] but if it is found that confidence and accuracy are aligned, lower levels of confidence could cue physicians to deliberate further or seek additional diagnostic help.

Empirical evidence supporting the notion that physicians are overconfident in their diagnostic accuracy is scarce.[7-13] The few studies that have examined the relationship between diagnostic accuracy and confidence have produced mixed results. Moreover, these studies were limited in scope: they focused on only 1 component of the diagnostic process (largely diagnostic test interpretation),[7,10,11,13] included only trainees,[7,11] and/or involved chart reviews or other proxy measures of confidence that might not reliably reflect physicians' confidence.[8,9,12] Because overconfidence or underconfidence might change with evolution of diagnosis (ie, through history, examination, and testing), it is essential to use more reliable measures of confidence and to gather confidence ratings throughout the evolution.

It is also unclear how confidence relates to physicians' resource requests, such as specialized tests, referrals, second opinions, and use of reference materials to facilitate diagnosis. If physicians' accuracy and confidence are miscalibrated, their decisions to investigate further could be misguided, leading to delay of diagnosis, harm, or inappropriate resource utilization.

To investigate the relationship between diagnostic accuracy and confidence, we evaluated diagnostic calibration (ie, the relationship between diagnostic accuracy and confidence) during evolution of the diagnostic process. We hypothesized that calibration would improve with evolution (ie, as additional diagnostic data became available), but would remain unchanged across cases of varying difficulty (ie, accuracy and confidence would decrease with more difficult cases, but calibration would remain unchanged). As a secondary objective, we explored the relationship between confidence and additional resource utilization to facilitate diagnosis.

Methods

Participants and Setting

We recruited general internal medicine physicians from QuantiaMD.com, an online educational community of physicians where physicians can review clinical evidence from experts and exchange knowledge. Some physicians also receive continuing medical education credit for participating in certain educational activities at this site, although we did not provide such credit or any other incentives for participation. The study protocol was approved by our local institutional review board.

Materials

Participants assessed 4 clinical vignettes based on real-world cases, authored by medical experts. Cases were selected from a pool of validated cases used in previous studies.[8,9,14-17] To select cases of relatively different difficulty levels, we took into account difficulty ratings previously assigned by 3 case authors, all experienced internists (Cronbach α interrater reliability, 0.83).[9,14] To determine if physicians adjust their confidence with case difficulty, we selected 2 relatively *easier* cases (ratings of 3.2 and 3.7 of 7, with 7 indicating the highest level of difficulty) and 2 *more difficult* cases (ratings of 6.0 and 5.2). (The Case Synopses subsection herein summarizes case characteristics.) We avoided cases described by the raters as rare diseases with atypical presentations or as missing key diagnostic data. We collected data according to 4 phases of case evolution: (1) chief complaint and medical history, (2) physical examination, (3) general laboratory and imaging, and (4) definitive or specialized laboratory and imaging. We developed questions to assess accuracy and self-reported confidence during the diagnostic process and pilot tested them with 6 internists using a paper-based assessment format. **Figure 1** depicts the procedure and questions asked in each case. We subsequently converted materials into a web-based format for presentation on QuantiaMD.com, including slides, visual text, and audio readout of text by a QuantiaMD.com narrator. The web-based format was pilot tested with additional physicians and a usability specialist who ensured the web-interface design was optimal.

Case Synopses

Easier Cases

Case 1 involved a 23-year-old Hispanic male migrant worker from Mexico who presented with right upper quadrant abdominal pain. In Mexico, he drank water from streams. Imaging at an outside hospital revealed 2 cystic liver lesions. His examination showed decreased breath sounds at the right lung base; laboratory findings included elevated alkaline phosphatase levels; and chest radiography showed right-sided effusion.
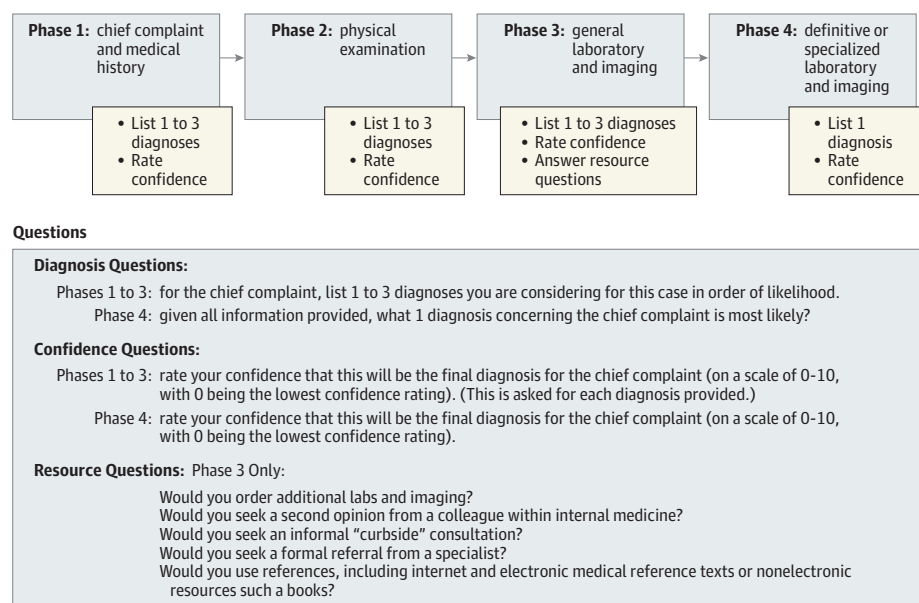
Case 2 involved a 60-year-old man with crampy lower abdominal pain of 3 weeks' duration, anemia, and recent weight loss. His examination showed pale conjunctivae and palpable liver edge 2 cm below the right costal margin. Results of occult blood testing were positive. His laboratory findings revealed a hemoglobin level of 5g/dL with mean corpuscular volume (MCV) of 55 fL.

More Difficult Cases

Case 3 involved a 25-year-old woman with headaches, macular rash, and vertigo. Her medical history included genital herpes and chlamydia. Her physical examination revealed a faint erythematous macular rash on her forearms that did not involve the palms. The findings of her neurologic examination were nonfocal. Her laboratory results revealed an erythrocyte sedimentation rate (ESR) of 77 mm/h.

Case 4 involved a 68-year-old African American man who presented with fever, fatigue, and arthralgias with wrist and shoulder pain that began 4 weeks earlier. His examination find-

Figure 1. Study Procedure and Questions Asked in the 4 Diagnostic Phases for Each Case

Phase 1: chief complaint and medical history
- List 1 to 3 diagnoses
- Rate confidence

Phase 2: physical examination
- List 1 to 3 diagnoses
- Rate confidence

Phase 3: general laboratory and imaging
- List 1 to 3 diagnoses
- Rate confidence
- Answer resource questions

Phase 4: definitive or specialized laboratory and imaging
- List 1 diagnosis
- Rate confidence

Questions

Diagnosis Questions:

Phases 1 to 3: for the chief complaint, list 1 to 3 diagnoses you are considering for this case in order of likelihood.

Phase 4: given all information provided, what 1 diagnosis concerning the chief complaint is most likely?

Confidence Questions:

Phases 1 to 3: rate your confidence that this will be the final diagnosis for the chief complaint (on a scale of 0-10, with 0 being the lowest confidence rating). (This is asked for each diagnosis provided.)

Phase 4: rate your confidence that this will be the final diagnosis for the chief complaint (on a scale of 0-10, with 0 being the lowest confidence rating).

Resource Questions: Phase 3 Only:

Would you order additional labs and imaging?
Would you seek a second opinion from a colleague within internal medicine?
Would you seek an informal "curbside" consultation?
Would you seek a formal referral from a specialist?
Would you use references, including internet and electronic medical reference texts or nonelectronic resources such a books?

ings were significant for a temperature of 38.9°C, heart rate of 120 bpm, and pale conjunctivae. He had no joint swelling but had pain in his shoulders on abduction to 90°. Hemoglobin level was 9.3 g/dL; MCV, 73 fL; white blood cell count, 9.2 × 10³/μL; creatine kinase, 12 U/L; ESR, 100 mm/h. Results of iron studies and chest radiography were normal.

Procedure

To obtain participation from a broad physician audience, we recruited physicians through e-mails with support from staff from QuantiaMD.com. QuantiaMD.com uses the National Provider Identifier database to verify "physician status" during account creation, at which time physicians record their specialty. Participants were informed that the study would explore the relationship between physicians' diagnostic accuracy and confidence and the relationship between accuracy and resource utilization. Consent was implied if practitioners responded. After consenting, physicians provided demographic information including age, sex, race, years of medical experience since residency, birth country, country of medical education, and work environment. They were then given instructions on case assessment and completed a different practice case.

The 4 vignettes were presented in random order. In each case, physicians used free text to record on the web application 1 to 3 differential diagnoses after each of the first 3 successive phases of the diagnostic process and provided judgments of confidence for each diagnosis. Only 1 "final" diagnosis and confidence judgment was recorded after the fourth phase. Before being shown definitive diagnostic data (phase 4), participants were asked to note which additional resources they would request to make a definitive diagnosis. These included additional diagnostic tests, second opinions, curbside consultations, referrals, and reference materials such as electronic

and nonelectronic books. All participants were shown the same definitive data in phase 4, regardless of what they requested. This data collection strategy ensured that calibration information was gathered throughout the diagnostic process, simulating information discovery seen in clinical practice and allowing for comparison of physicians. The procedure took approximately 1 hour to complete.

We used a repeated-measures study design, in which accuracy and confidence were collected repeatedly from each participant and were evaluated at each phase of the diagnostic process and with varied case difficulty.

Outcome Measures

Accuracy was scored by comparing participants' free-text responses to the true diagnoses provided by the case authors. The list of acceptable answers, including acceptable variations of a diagnosis name, was compiled and provided by the original case authors, 3 expert internists. One of our physician team members (D.M.) further verified these responses. Spelling errors did not count against correctness. Each diagnosis provided was scored as 1 (agreement with true diagnosis) or 0 (disagreement with true diagnosis). For each of phases 1 through 3, participants were given credit for correctly diagnosing the case if the true diagnosis was included in their differential diagnoses. For phase 4, credit was given when the final diagnosis given was the true diagnosis. All phases were scored independently.

Confidence data were rated using an 11-point scale (0-10; with the anchors of 0, indicating lowest confidence rating, and 10, highest confidence rating). The confidence value used for data analysis corresponded to the value associated with the participant's correct diagnosis or to the top diagnosis listed if none of the diagnoses provided were correct. Physicians were asked to list diagnoses in order of likelihood, so the top diag-

jamainternalmedicine.com

nosis coincided with their highest confidence level. Confidence responses were rescaled (from 0-10 to 0-1) by dividing by 10, so confidence and accuracy would span the same numbers.

We used methods from cognitive psychology to compare a person's accuracy with their perception of confidence in their accuracy and thus provide information in terms of calibration or alignment between these 2 concepts and the direction of miscalibration.[18] We assessed how accuracy is aligned with confidence, ranging from 0 (best possible alignment) to 1 (worst possible alignment). Calibration is calculated as "the weighted mean of the squared difference between confidence and proportion correct for each confidence level." We also used the over-under index (O-U index), which denotes direction and magnitude of miscalibration, ranging from −1 (highest possible level of underconfidence) to +1 (highest possible level of overconfidence). Miscalibration is computed as the difference between confidence and accuracy.[18]

Resource requests were treated dichotomously and scored as 0 (No) or 1 (Yes) for whether participants indicated they would request each additional resource.

Statistical Analysis

Diagnostic accuracy, confidence, and calibration (both alignment and miscalibration direction) were assessed using repeated-measures analysis of variance as a function of case difficulty and diagnostic phase. Both were treated as repeated measures. Average accuracy was also assessed as a function of physician experience, sex, country of medical education, and work environment using Pearson (for experience) and point-biserial correlation analyses (for all other variables).

Resource requesting was analyzed using repeated-measures logistic regression using case vignette as the repeated measure. This was only analyzed for phase 3 data, where this question was asked. Predictors included confidence in diagnostic accuracy, diagnostic accuracy (incorrect vs correct), case difficulty (easier vs more difficult), years of experience, and country of medical education (United States vs international). Each resource—additional diagnostic tests, second opinions, curbside consultations, referrals, and reference materials—was regressed onto these predictors separately.

Results

Of 658 physicians who viewed the e-mail invitation containing study details, 118 completed the study (17.9% recruitment rate). Physicians represented at least 33 US states and territories and a wide range of clinical experience and home institutions (**Table 1**). Slightly over one-third (34.7%) were international medical graduates, consistent with their representation among US internal medicine physicians nationally (37%).[19] Sex and race demographics also coincided with the US internal medicine physician workforce.[20]

Diagnostic Accuracy

Figure 2 illustrates changes in diagnostic accuracy as the diagnostic process evolved in both easier and more difficult cases.

Table 1. Characteristics of Physician Participants

| Demographic Characteristic	Participants[a]
Age, mean (range), y	48.4 (25-85)
Experience (years since residency), mean (range), y	16.5 (0-56)
Sex	
Male	85 (72.0)
Female	33 (28.0)
Race	
White	76 (64.4)
Asian	35 (29.7)
Black or African American	4 (3.4)
Other	3 (2.5)
Country of birth	
United States	71 (60.2)
Other	47 (39.8)
Country of medical education	
United States	77 (65.3)
International	41 (34.7)
Type of work institution	
Private	99 (83.9)
Government	19 (16.1)
Academic affiliation of work institution	
Nonacademic	72 (61.0)
Academic	46 (39.0)

[a] Unless otherwise indicated, data are number (percentage) of physician participants.

Compared with easier cases, diagnostic accuracy was worse for more difficult cases (55.3% vs 5.8% accuracy; $F_{1,117} = 231$; $\eta_p^2 = 0.66$) ($P < .001$), consistent with our expectations. Neither the 2 easier cases nor the 2 more difficult cases differed significantly from each other on any other measure (confidence, calibration, or O-U index). For subsequent analyses, we thus grouped together the 2 easier cases into the easier category and the 2 more difficult cases into the more difficult category. Accuracy remained relatively stable during the first 3 phases of the diagnostic process but declined slightly in the last phase, when participants specified the final diagnoses ($F_{3,351} = 14.9$; $\eta_p^2 = 0.11$) ($P < .001$) (Figure 2). Average diagnostic accuracy did not vary significantly with physician experience ($r_{118} = 0.08$; $P = .39$), sex ($r_{118} = 0.02$; $P = .85$), country of medical education ($r_{118} = −0.11$; $P = .24$), or work environment (government vs private [$r_{118} = 0.09$; $P = .35$]; academic vs nonacademic [$r_{118} = −0.07$; $P = .43$]).

Confidence

As shown in Figure 2, physicians' confidence in their accuracy was only slightly lower with more difficult cases ($F_{1,117} = 37.4$, $\eta_p^2 = 0.24$) ($P < .001$); participants gave an average confidence rating of 7.2 for easier cases and 6.4 for more difficult cases (on a scale of 0-10). Confidence varied slightly across diagnostic phases, such that confidence was highest after general diagnostic test data were given but before any definitive or specialized diagnostic data were given (mean confidence levels for diagnostic phases 1, 2, 3, and 4 were 6.7, 6.7, 7.0, and 6.8, respectively [$F_{3,351} = 9.94$, $\eta_p^2 = 0.08$] [$P < .001$]). Average confi-

Figure 2. Physicians' Mean Diagnostic Accuracy and Confidence in That Accuracy as a Function of Diagnostic Phase and Case Difficulty (Easier vs More Difficult)

Lab indicates laboratory testing; error bars represent ±1 SEM.
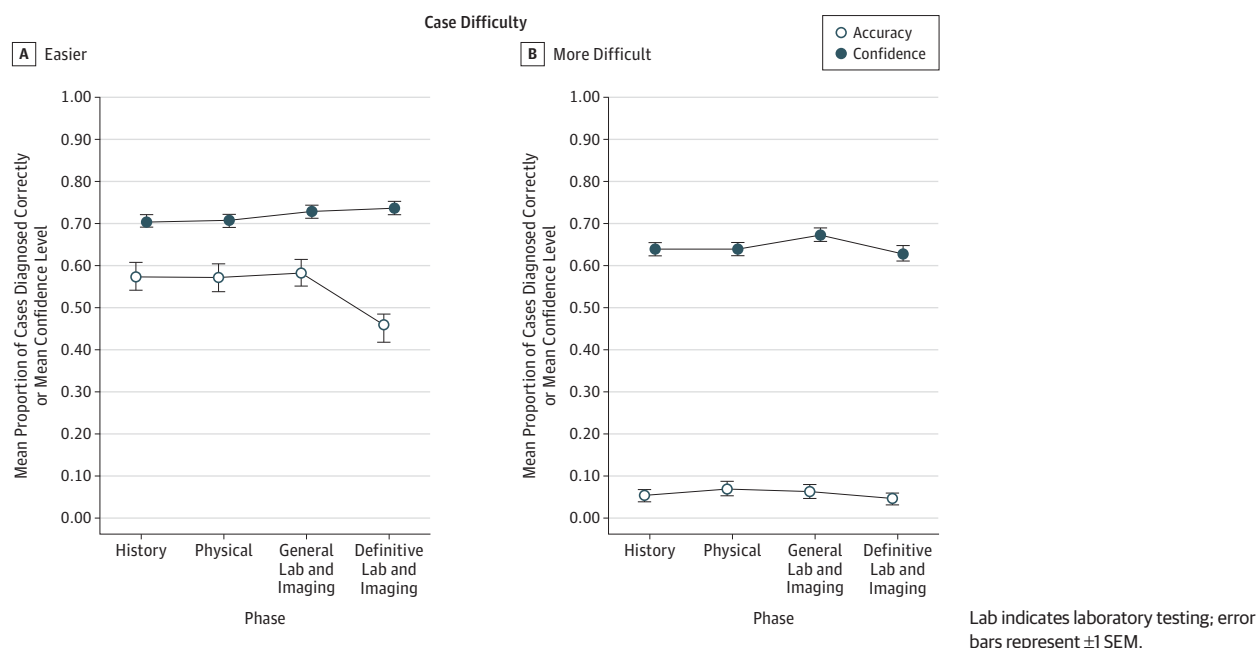
dence did not vary significantly with physician experience ($r_{118} = -0.06$; $P = .53$). However, within easy cases and within difficult cases, confidence levels were highly variable within individual physicians (see the eFigure in the Supplement).

Diagnostic Calibration

Alignment (Calibration)
According to our hypothesis, both accuracy and confidence should change between easier and more difficult cases, such that calibration remains unchanged. However, results show that calibration (ranging from 0 [best possible alignment] to 1 [worst possible alignment]) deteriorated with increasing case difficulty, with a calibration of 0.31 for easier cases vs 0.45 for more difficult cases ($F_{1,117} = 37.8$; $\eta_p^2 = 0.24$) ($P < .001$), suggesting that accuracy and confidence were less aligned for more difficult cases. This was because accuracy decreased substantially for more difficult cases, but confidence did not decrease in a complementary fashion (as shown in the large gap in Figure 2B). Furthermore, calibration worsened slightly during the diagnostic phases 3 and 4 ($F_{3,351} = 6.27$; $\eta_p^2 = 0.05$) ($P < .001$). For easier cases, calibration was least aligned in the fourth phase ($P = .001$), whereas for more difficult cases, it was least aligned in the third phase ($P = .001$). This indicates that accuracy and confidence did not change proportionally as the diagnostic process evolved, contrary to our hypothesis that calibration would improve as additional data became available in the diagnostic process.

Direction and Magnitude of Miscalibration (O-U Index)
Miscalibration consistently occurred in the overconfident direction (mean O-U index, 0.38 on a scale of −1 [representing highest possible level of underconfidence] to +1 [representing highest possible level of overconfidence]). The magni-

tude of overconfidence increased with case difficulty ($F_{1,117} = 169$; $\eta_p^2 = 0.59$) ($P < .001$); mean O-U index of 0.17 for easier cases vs 0.59 for more difficult cases. Additionally, overconfidence was slightly higher in the fourth diagnostic phase (mean of 0.43) ($P < .001$), an effect that was seen in easier cases ($P < .001$) but not for more difficult ones ($P = .11$). In difficult cases, overconfidence remained at a high level throughout the diagnostic process.

Resource Requests
As summarized in **Table 2**, increased confidence was related to decreased likelihood of requesting additional diagnostic tests (and vice versa) ($\chi_1^2 = 7.06$; odds ratio [OR], 0.91) ($P = .01$). Higher case difficulty was related to resource requests for only 1 type of resource: an increased likelihood of requesting additional reference materials ($\chi_1^2 = 13.5$; OR, 1.83) ($P = .01$). Diagnostic accuracy was not significantly related to the request of any type of resource.

Physician characteristics related to resource requests included years of experience and country of medical education. Specifically, increased experience was associated with decreased likelihood to request second opinions ($\chi_1^2 = 4.38$; OR, 0.98) ($P = .04$), curbside consultations ($\chi_1^2 = 9.98$; OR, 0.96) ($P = .002$), and reference materials ($\chi_1^2 = 6.41$; OR, 0.97) ($P = .02$). Having graduated from a medical school outside the United States was associated with increased likelihood of requesting reference materials ($\chi_1^2 = 5.75$; OR, 3.15) ($P = .02$) (Table 2).

Discussion
Using a case vignette–based study, we evaluated the relationship between diagnostic accuracy and confidence in that ac-

Table 2. Relationship Between Resource Requests and Confidence, Accuracy, Case Difficulty, and Health Care Provider Characteristics

Resource	Odds Ratio[a] (P Value)				
	Confidence	Diagnostic Accuracy (*Incorrect*[b] vs Correct)	Case Difficulty (*Easier*[b] vs More Difficult)	Experience	Medical School (*United States*[b] vs International)
Additional diagnostic tests	0.91 (.01)	0.91 (.55)	0.82 (.09)	0.98 (.13)	0.93 (.82)
Second opinions	0.97 (.43)	0.71 (.09)	0.94 (.68)	0.98 (.04)	0.61 (.11)
Curbside consultations	0.91 (.09)	0.90 (.56)	1.07 (.61)	0.96 (.002)	0.73 (.42)
Referrals	0.88 (.13)	1.16 (.65)	0.79 (.39)	1.02 (.18)	1.48 (.26)
References	1.01 (.91)	0.69 (.17)	1.83 (.01)	0.97 (.02)	3.15 (.02)

[a] Odds ratio represents change in odds of requesting a resource (eg, in comparison with obtaining one's medical education in the *United States,* being educated internationally resulted in a 3.15 increase in odds of requesting reference materials, or being 3.15 times more likely to request reference materials).

[b] Italics denote reference categories.

curacy as diagnosis evolved and as cases varied in difficulty. As expected, we found that diagnostic accuracy decreased when physicians were faced with more difficult cases. However, physicians' confidence decreased only slightly with difficult cases, and the amount it decreased was not practically or clinically meaningful, especially in view of a large reduction in accuracy. When we evaluated the relationship between confidence and accuracy more directly using calibration measures, we found these 2 to be less aligned when physicians were presented with more difficult cases (ie, their diagnostic calibration worsened). The level of overconfidence also increased with the difficult cases, contrary to our hypothesis that both accuracy and confidence would decrease when physicians were faced with difficult cases. Furthermore, while confidence increased as cases evolved and as more data became available, accuracy did not; this also resulted in increasing overconfidence. This was again contrary to our hypothesis that physicians would become better calibrated as more information became available in the diagnostic process.

Our study has several strengths. To our knowledge, this is the first study to directly examine the relationship between physicians' diagnostic accuracy and confidence within the context of the evolution of the diagnostic process. We directly asked physicians to rate their confidence rather than rely on proxy measures of confidence used in other work.[8,9,12] While self-report confidence measures may not reflect physicians' actual confidence, they are more likely to be accurate than when external raters attempt to infer physician confidence from medical record reviews, a technique used in other studies. We also assessed accuracy and confidence of a wide range of practicing physicians whose demographics appear representative of practicing US internal medicine physicians.

The overall diagnostic accuracy was rather low—31% across the 4 cases. However, a previous study[14] using these 4 as well as many other cases revealed an average accuracy only moderately higher at 43%; average difficulty ratings were comparable. The previous study included only academic physicians who might have performed better on difficult cases because they tend to see more complex cases while practicing in tertiary medical centers. For our study purposes, using cases that could test whether physicians adjust their confidence when faced with difficult cases was important. Thus, we deliberately chose 2 difficult cases to test our hypothesis, although

we acknowledge that it is difficult to know what "average" case difficulty is. In general, we expected that regardless of performance, confidence level would ideally reflect physician's accuracy. These findings might imply that overconfidence, which increased with both case difficulty and as diagnostic processes evolved, might impede recognition of faulty reasoning. Our results are similar to what has been found in other domains including logical reasoning, such as when confidence—but not accuracy—increases with more time to think about whether a conclusion based on 2 assertions is valid.[21]

Physician confidence was also related to how often they requested an important additional resource; diagnostic tests were requested less often when confidence level was higher (regardless of whether that confidence was accurately placed). When faced with more difficult cases, the only resource physicians were more likely to request was reference materials. In essence, physicians did not request more second opinions, curbside consultations, or referrals in situations of decreased confidence, decreased accuracy, or when diagnosing difficult cases. These findings suggest that physicians might not request the required additional resources to facilitate diagnosis when they most need it.

We also found that physician characteristics were related to requests for additional resources. Specifically, increased experience was associated with decreased likelihood of requesting second opinions, curbside consultations, and reference materials, regardless of diagnostic accuracy. This corresponds with findings in previous research, suggesting that there might be differences in how experienced physicians approach the diagnostic process.[22] Also, being an international medical graduate was associated with increased likelihood of requesting reference materials, suggesting that medical school training, or earlier training, might influence the way physicians seek help in making diagnoses. Future research should further examine how confidence is related to the evolving reasoning process and how this relates to physicians' risk aversion.[23]

Our findings could have important implications for continuing physician education, where little emphasis is placed on diagnostic calibration. If confirmed in real clinical settings, our findings of a weak association between diagnostic accuracy and confidence provide evidence needed to develop interventions to improve physicians' diagnostic calibration and resource use as well as related diagnostic outcomes. For example, using feedback for recalibration has been sug-

gested and may be a necessary step for improving diagnostic reasoning.[24] An additional solution might include requiring physicians to better justify diagnoses: justification of answers, particularly revolving around disconfirming evidence, has been found to decrease overconfidence in other domains.[25,26] Engaging patients in creative ways during the patient-provider encounter might provide a way to circumnavigate overconfidence because diagnostic errors have a predilection for this part of the diagnostic process.[1] Debiasing techniques (cognitive strategies used to overcome one's biases in judgment and thinking) and reflective practice (critical self-deliberation about one's own decision making) might also be useful[27,28]; however, future research should evaluate the efficacy of these techniques to improve diagnostic calibration.

Our study has several limitations. Our methods of case delivery limit real-world validity of the study. However, we provided participants with diagnostic information and requested differential diagnosis to simulate real-world diagnostic decision making, which is otherwise difficult to elicit. Additionally, the artificial diagnostic environment allowed us to control case selection and require all participants to assess the same cases, a difficult scenario to achieve in real-world clinical settings. While cases were assessed over the web, they were based on real-world cases, and their presentation simulated the natural evolution of diagnosis. To give busy physicians flexibility in completing the cases, we permitted them to complete cases in different sessions and thus were unable to control whether they completed all cases at once. However, it is unlikely that participants' knowledge base or pattern of responding to confidence judgments changed between sessions. Moreover, the order in which the participants completed cases was randomized, lessening the potential adverse effects of this flexibility. Many factors (access, financial status, and other factors) other than confidence and/or accuracy may affect resource utilization, but our simulated setting should have minimized their impact. Additionally, while categorization of the accuracy of diagnosis relied on manual coding of free-text responses and hence may be subject to bias, we used acceptable variations of diagnosis provided by original case authors to minimize it. Another limitation is that physicians may have been unwilling to disclose uncertainty, as has been suggested by others.[29] While this is plausible, our study conditions did not exert any social or legal pressures that might impel physicians to maintain a pretense of certainty. Finally, because diagnostic behaviors and cognitive skills of typical practicing physicians are neither well understood nor well known, our study population might not be representative of practicing internists. However, demographics of the physicians do mirror the general physician population.

In conclusion, our study suggests that the association between physicians' diagnostic accuracy and their confidence in that accuracy may be poor and that physicians may not request the required additional resources to facilitate diagnosis when they most need it. These mismatched associations might prevent physicians from reexamining difficult cases when their diagnosis is incorrect. Improving these associations and the use of potential resources in handling difficult cases could potentially reduce diagnostic error.

REFERENCES

1. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med.* 2013;173(6):418-425.

2. Graber M. Diagnostic errors in medicine: a case of neglect. *Jt Comm J Qual Patient Saf.* 2005;31(2):106-113.

3. Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. *Qual Saf Health Care.* 2006;15(3):159-164.

4. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med.* 2009;169(20):1881-1887.

5. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(5)(suppl):S2-S23.

6. Graber ML, Berner ES, Suppl eds. *Diagnostic Error: Is Overconfidence the Problem?* http://www.amjmed.com/issues?issue_key= S0002-9343%2808%29X0007-5. Accessed August 8, 2013.

7. Davis DP, Campbell CJ, Poste JC, Ma G. The association between operator confidence and accuracy of ultrasonography performed by novice emergency physicians. *J Emerg Med.* 2005;29(3):259-264.

8. 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. *Stud Health Technol Inform.* 2001;84(Pt 1):454-458.

9. Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? implications for decision support and error reduction. *J Gen Intern Med.* 2005;20(4):334-339.

10. Haiyan H, Min D. Ultrasound operators' confidence influences diagnosis of ovarian tumors:

a study in China. *Asian Pac J Cancer Prev*. 2011;12(5):1275-1277.

11. Mann D. The relationship between diagnostic accuracy and confidence in medical students. http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED358110. Accessed July 16, 2013.

12. Podbregar M, Voga G, Krivec B, Skale R, Pareznik R, Gabrscek L. Should we confirm our clinical diagnostic certainty by autopsies? *Intensive Care Med*. 2001;27(11):1750-1755.

13. Yazbek J, Ameye L, Testa AC, et al. Confidence of expert ultrasound operators in making a diagnosis of adnexal tumor: effect on diagnostic accuracy and interobserver agreement. *Ultrasound Obstet Gynecol*. 2010;35(1):89-93.

14. Elstein AS, Friedman CP, Wolf FM, et al. Effects of a decision support system on the diagnostic accuracy of users: a preliminary report. *J Am Med Inform Assoc*. 1996;3(6):422-428.

15. Friedman CP, Elstein AS, Wolf FM, et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. *JAMA*. 1999;282(19):1851-1856.

16. Murphy GC, Friedman CP, Elstein AS, Wolf FM, Miller T, Miller JG. The influence of a decision

support system on the differential diagnosis of medical practitioners at three levels of training. *Proc AMIA Annu Fall Symp*. 1996:219-223.

17. Wolf FM, Friedman CP, Elstein AS, et al. Changes in diagnostic decision-making after a computerized decision support consultation based on perceptions of need and helpfulness: a preliminary report. *Proc AMIA Annu Fall Symp*. 1997:263-267.

18. Weber N, Brewer N. Confidence-accuracy calibration in absolute and relative face recognition judgments. *J Exp Psychol Appl*. 2004;10(3):156-172.

19. American Medical Association International Medical Graduates Section Governing Council. *International Medical Graduates in American Medicine: Contemporary Challenges and Opportunities.* Chicago, IL: American Medical Association; 2013.

20. US Dept of Health and Human Services. *The Physician Workforce: Projections and Research Into Current Issues Affecting Supply and Demand.* Rockville, MD: Health Resources and Services Administration; 2008.

21. Shynkaruk JM, Thompson VA. Confidence and accuracy in deductive reasoning. *Mem Cognit*. 2006;34(3):619-632.

22. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med*. 1990;65(10):611-621.

23. Pines JM, Hollander JE, Isserman JA, et al. The association between physician risk tolerance and imaging use in abdominal pain. *Am J Emerg Med*. 2009;27(5):552-557.

24. Schiff GD. Minimizing diagnostic error: the importance of follow-up and feedback. *Am J Med*. 2008;121(5)(suppl):S38-S42.

25. Arkes HR, Christensen C, Lai C, Blumer C. Two methods of reducing overconfidence. *Organ Behav Hum Dec*. 1987;39(1):133-144.

26. Koriat A, Lichtenstein S, Fischhoff B. Reasons for confidence. *J Exp Psychol Hum Learn Mem*. 1980;6(2):107-118.

27. Croskerry P, Norman G. Overconfidence in clinical decision making. *Am J Med*. 2008;121(5)(Suppl):S24-S29.

28. Mamede S, Schmidt HG, Rikers R. Diagnostic errors and reflective practice in medicine. *J Eval Clin Pract*. 2007;13(1):138-145.

29. Katz J. Why doctors don't disclose uncertainty. *Hastings Cent Rep*. 1984;14(1):35-44.

--- **Invited Commentary** ---

Known Unknowns and Unknown Unknowns at the Point of Care

Gurpreet Dhaliwal, MD

Lifelong learning (LLL) in medicine is at once a value, a set of skills, and a social pact. It is part of our professional identity, a core competency of the modern clinician, and an inviolable obligation to our patients. Despite its alignment with professional ideals, LLL remains a voluntary and self-motivated pursuit with no direct extrinsic rewards, which raises questions about what can be done to enhance its value and facilitate its widespread enactment.

Clinicians consistently report that patient problems from their own practices are their greatest motivators for learning, yet most of their questions go unanswered.[1]

Health systems have an opportunity to capitalize on this motivation and unmet need by optimizing point-of-care learning (POCL), where questions triggered by patient care activities are answered in real time. All physicians know, however, that this is easier said than done.

In this issue of *JAMA Internal Medicine*, Cook and colleagues[2] take a granular look at what stands in the way of on-the-job POCL. Based on interviews with 50 physicians, the authors model the multiple decisions clinicians face when contemplating information searching while caring for a patient, including whether to search, when to search, where to search, and when to stop. It will come as no surprise that insufficient time was the greatest barrier to POCL (along with case complexity and information overload) and that efficiency was the most important determinant in selecting an information source. These findings match those of a recent study of POC resource selection by internal medicine residents, who prioritized efficiency and credibility of sources (eg, UpToDate and Google) and cited time and information overload as their principal barriers. Both experienced physicians and residents share a common goal of learning by finding "the most trusted and relevant information in the shortest time."[3(p790)]

The first node in the model described by Cook et al[2] (whether to initiate POCL) may be more complicated than meets the eye. While most physicians would claim to be capable of determining when they need assistance, studies consistently reveal that physicians' ability to self-assess knowledge and skills is inaccurate and that we maintain blind spots to our deficiencies.[4] This misperception may arise in part from the universal human tendency to mistake confidence for competence. If confidence and skill were tightly calibrated, physicians could rely on their confidence as a guide for when they