

Brief Communication

Impact of digitally acquired peer diagnostic input on diagnostic confidence in outpatient cases: A pragmatic randomized trial

Elaine C. Khoong ^{1,2}, Valy Fontil ^{1,2}, Natalie A. Rivadeneira^{1,2}, Mekhala Hoskote³, Shantanu Nundy^{4,5}, Courtney R. Lyles ^{1,2}, and Urmimala Sarkar^{1,2}

¹Division of General Internal Medicine, Zuckerberg San Francisco General Hospital, Department of Medicine, University of California, San Francisco, San Francisco, California, USA, ²Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, California, USA, ³Berkeley School of Public Health and UCSF School of Medicine, University of California, Berkeley–University of California, San Francisco Joint Medical Program, Berkeley, California, USA, ⁴Milken Institute School of Public Health, Department of Health Policy and Management, George Washington University, Washington, DC, USA and ⁵Accolade, Inc. Plymouth Meeting, Pennsylvania, USA

Corresponding Author: Elaine C. Khoong, MD, MS, Division of General Internal Medicine at Zuckerberg San Francisco General Hospital, Building 10, Ward 13, 1001 Potrero Avenue, San Francisco, CA 94110, USA; elaine.khoong@ucsf.edu

Received 19 July 2020; Editorial Decision 22 October 2020; Received 19 July 2020

ABSTRACT

Objective: The study sought to evaluate if peer input on outpatient cases impacted diagnostic confidence.

Materials and Methods: This randomized trial of a peer input intervention occurred among 28 clinicians with case-level randomization. Encounters with diagnostic uncertainty were entered onto a digital platform to collect input from ≥ 5 clinicians. The primary outcome was diagnostic confidence. We used mixed-effects logistic regression analyses to assess for intervention impact on diagnostic confidence.

Results: Among the 509 cases (255 control; 254 intervention), the intervention did not impact confidence (odds ratio [OR], 1.46; 95% confidence interval [CI], 0.999–2.12), but after adjusting for clinician and case traits, the intervention was associated with higher confidence (OR, 1.53; 95% CI, 1.01–2.32). The intervention impact was greater in cases with high uncertainty (OR, 3.23; 95% CI, 1.09–9.52).

Conclusions: Peer input increased diagnostic confidence primarily in high-uncertainty cases, consistent with findings that clinicians desire input primarily in cases with continued uncertainty.

Key words: information technology, clinical decision making, computer-assisted medical decision making, diagnostic confidence, pragmatic clinical trial, peer review

INTRODUCTION

Each year, $\sim 5\%$ of American adults experience a diagnostic error during outpatient care.¹ Diagnostic confidence impacts diagnostic error.^{2–5} Overconfidence reduces diagnoses considered and diagnostic assistance or tests pursued^{4,5}; underconfidence increases use of unnecessary diagnostic tests.^{5–7}

Input on clinical decision making, particularly early in the diagnostic process, impacts diagnostic confidence and accuracy.^{2,7–11} However, most studies on diagnostic confidence have utilized clinical vignettes and inpatient diagnoses.¹² To our knowledge, no studies have evaluated informatics interventions to impact diagnostic confidence in real-time outpatient clinical practice.

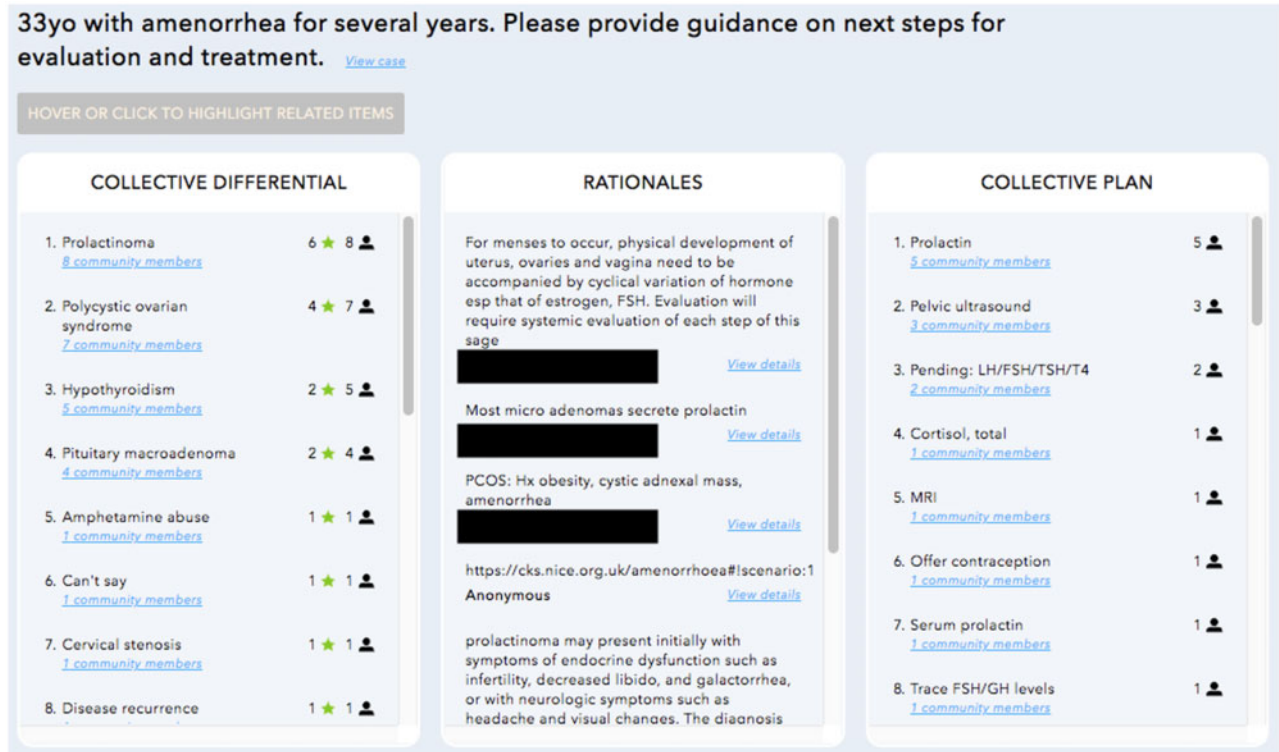


Figure 1. Example collective opinion from digital platform.

To evaluate the impact of peer input on outpatient cases with diagnostic uncertainty, we conducted a trial that utilized a digital tool to solicit and provide peer input on real cases. With the development of tools that can extract electronic health record clinical data through interoperable data standards, it is now more possible to use informatics tools to acquire peer input from external sources.^{13,14} In this report, we describe the rates of diagnostic confidence in outpatient encounters and the impact of peer input on confidence. We hypothesized that peer input would increase diagnostic confidence.

MATERIALS AND METHODS

The protocol for this pragmatic randomized trial¹⁵ of a diagnostic intervention was published¹⁶ and approved by our institutional review board. It is described subsequently and depicted in [Supplementary Figure 1](#).

Participants

Participants were primary care providers (PCPs) in 2 health systems: a safety net health system and an academic medical system. All PCPs (physicians or nurse practitioners) were eligible and recruited from May to November in 2018.

Case inclusion criteria

Physician scribes reviewed completed notes from PCPs' encounters and used criteria from the literature^{3,17} to identify encounters with potential for diagnostic uncertainty if they met 1 of 5 inclusion criteria: (1) new symptom or test abnormality, (2) unresolved symptom or test abnormality without definitive etiology, (3) test ordered to assess an unresolved concern, (4) empiric treatment documented, or (5) specialist referral for diagnostic assistance.

Intervention

The intervention was provision of peer opinions on the differential diagnosis and diagnostic steps. Physician scribes, who were primary care clinicians in the same healthcare system as participating PCPs, entered 1-line summaries with relevant history, exam, and tests onto a digital platform (the Human Diagnosis Project [Human Dx]; Human Dx, San Francisco, CA) ([Supplementary Figure 2](#)). This platform allows clinicians to submit a clinical case to elicit feedback on the diagnosis and plan as well as to provide feedback on submitted cases. Using Human Dx, we solicited opinions about diagnoses and next steps from an online community of U.S.-based, attending internists or family medicine physicians who had solved 1 Human Dx case in the last 3 months.

Once ≥ 5 clinicians responded, a link to the collective diagnostic opinion was emailed to PCPs. At the time of this study, the Human Dx used a proportionally weighted algorithm to create the collective diagnostic opinion based on individual user responses ([Supplementary Figure 3](#)). This has been described in detail previously,¹⁸ but in brief, the platform created a ranked list of diagnoses (collective differential) and recommendations (collective plan) ([Figure 1](#)). This ranked list was ordered to reflect how frequently the diagnosis or plan appears among individual user responses and its location on an ordered list (eg, top vs fifth diagnosis) in each user's response. All included cases were entered on the Human Dx platform, but PCPs only viewed the output for intervention cases. The collective opinion was viewed by PCPs a median of 11 days in the control group and 12 days in the intervention group after the encounter (interquartile range, 8–16 days for both groups; $P = .485$ for difference between median). The peer input was distributed at the same time a survey was administered to collect the outcomes and case perceptions described below. In comparison with the planned protocol, the study protocol was frequently delayed at several steps: PCPs completing

clinical notes 3+ days after encounters, scribes entering cases >1 day after note completion, collective opinion distributed to PCPs 5+ days after case entry, and PCPs viewing the input 3+ days after it was provided (workflow in [Supplementary Figure 1](#)). Collective opinions were solicited for all cases but viewed by PCPs only for intervention cases.

Randomization

We used a random number generator to block randomize (block size of 4) at the case level, stratified by clinician to ensure PCPs had similar number of cases in each arm. The research analyst responsible for randomization and allocation did not conduct data analysis or reveal assignment to scribes or clinician respondents who provided diagnostic input, who were all blinded to the intervention assignment.

Outcome

The primary outcome for this study and in this report is self-reported diagnostic confidence (not at all, somewhat, moderately, and very confident), which we dichotomized as low (not at all or somewhat) vs high (moderate or very) for the primary analyses. As described in the protocol, we chose this outcome based on (1) its known impact on medical management, (2) prior feedback from clinicians that impacting confidence increases utility of this tool, and (3) the likelihood that it is impacted by the intervention (in comparison with outcomes such as time to diagnosis).^{5,16,19,20} In prior literature, confidence was measured on an 11-point scale,⁵ but we modified the scale based on pilot testing with clinicians who reported ability to distinguish only 4 levels of confidence. Our protocol planned for our outcome to be dichotomized as “not at all” vs all other levels of confidence, but due to low rates of “not at all confident” responses (<10%) ([Supplementary Table 1](#)), we defined low confidence as not at all or somewhat confident. For this article, in intervention cases we also describe if PCPs self-reported that viewing the collective opinion influenced decision making or resulted in changes to plan. We collected both outcomes via a survey ([Supplementary Appendix](#)).

Covariates

PCP characteristics (professional degree [physician vs nurse practitioner], years in practice, clinic setting) and case perceptions (level of difficulty [not at all or somewhat vs moderate or high], diagnostic uncertainty [not at all or somewhat vs moderate or high]) were collected via PCP self-report in a survey. Patient characteristics (sex, race or ethnicity, age) were collected through chart review.

Survey procedure

After the collective opinion was collected, we distributed a survey to PCPs for each case. In the survey, PCPs first viewed a 1-line summary of the encounter then provided perceptions on case difficulty and diagnostic uncertainty (covariates). For cases in the control condition, PCPs then reported their diagnostic confidence. For intervention cases, PCPs first viewed the collective opinion then reported confidence and impact of the collective opinion on decision making or treatment plans.

Data analysis

We ensured balance in PCP characteristics, case perceptions, and patient characteristics between study arms by using bivariate mixed-effects logistic regressions to explore associations between the pri-

mary outcome and covariates as a single fixed predictor variable, with clinician as the cluster variable. For our primary analyses, we utilized a mixed-effects logistic regression model, clustered at the PCP level, with intervention status as the single fixed predictor variable of the binary outcome of diagnostic confidence. We also a priori planned to conduct a mixed-effects logistic regression analysis that adjusted for PCP characteristics, case perceptions, and patient characteristics, as we suspected that case-level randomization within PCPs may not have eliminated variability, given the relatively small number of PCPs. We conducted sensitivity analyses to assess the impact of covariates as well as interactions between case perceptions and PCP characteristics on the diagnostic confidence. In all models, neither patient characteristics nor PCP characteristics and case perception interactions were significant; inclusion of these variables also did not impact the significance of other variables, so the final adjusted model only included clinician characteristics and case perceptions. Given known differences in cognitive errors in cases with high vs low perceived uncertainty²¹ and the significant impact of baseline uncertainty in our adjusted analyses, we also conducted subgroup analyses within only the low-uncertainty cases and only the high-uncertainty cases.

RESULTS

We recruited 28 outpatient clinicians from May to November 2018. The majority were medical doctors (71%) with varying clinical experience ([Table 1](#)).

From these 28 clinicians, we identified 524 potential cases for study inclusion ([Figure 2](#)). We included 509 cases (255 control cases; 254 intervention cases) from encounters in August to December 2018, of which 127 (24%) were higher difficulty. The intervention and control cases were not significantly different in terms of clinician, case, or patient traits ([Table 2](#)).

The rate of high diagnostic confidence was 46% (n = 117 of 255) in control cases vs 54% (n = 136 of 254) in intervention cases. In the bivariate mixed-effects logistic regression with single fixed predictors, only low case difficulty and low diagnostic uncertainty (baseline case perceptions) were associated with high confidence ([Supplementary Table 2](#)). In the primary analysis, the intervention did not significantly impact confidence (odds ratio [OR], 1.46; 95% confidence interval [CI], 0.999-2.12, $P = .0504$) ([Figure 3](#)). However, in analysis controlling for clinician degree, clinician time in practice, and case perceptions, the intervention was associated with higher confidence (OR, 1.53; 95% CI, 1.01-2.32; $P = .04$) ([Supplementary Table 3](#)).

In a subgroup analysis stratified by baseline diagnostic uncertainty, the intervention was associated with confidence only in cases

Table 1. Characteristics of included primary care clinicians (n = 28)

Professional degree	
Nurse practitioner	8
Medical doctor	20
Years in practice	
Less than 5 y	10
5-9 y	5
10-20 y	8
More than 20 y	5
Health system	
Safety net clinic	22
Academic medical center	6

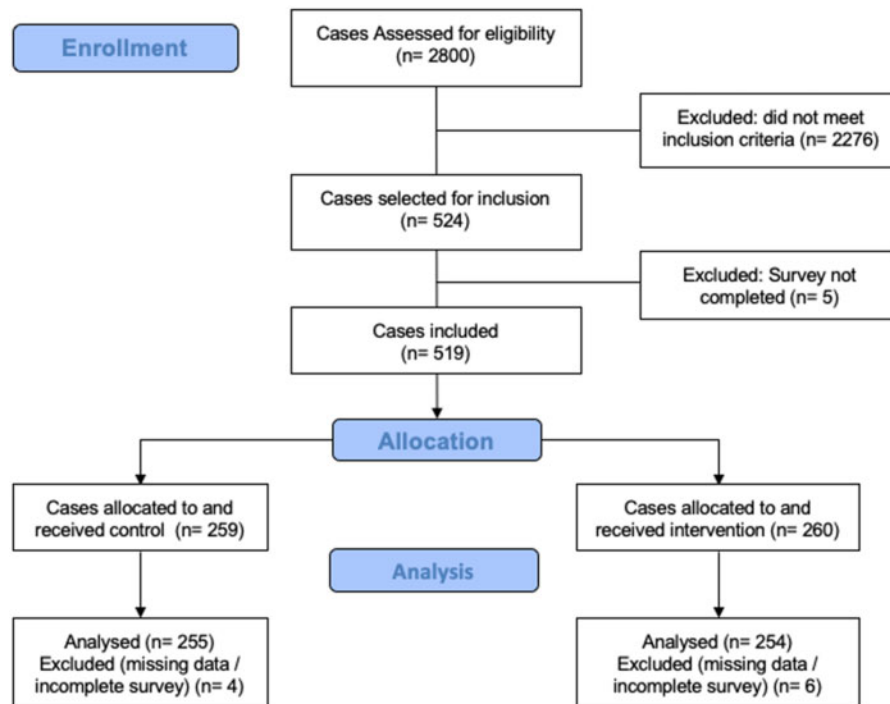


Figure 2. CONSORT diagram: case selection and inclusion.

Table 2. Characteristics of included cases

	Control (n = 255)	Intervention (n = 254)	P Value
Primary care clinician characteristics			
Nurse practitioner	35 (89)	37 (94)	.62
Years in practice			.78
Less than 5 y	39 (99)	41 (103)	
5-9 y	17 (43)	19 (49)	
10-20 y	31 (78)	28 (72)	
More than 20 y	14 (35)	12 (30)	
Case perceptions ^a			
Perceived higher difficulty	25 (64)	24 (62)	.86
Higher uncertainty in diagnosis	25 (65)	24 (62)	.78
Patient characteristics			
Female	59 (148)	55 (137)	.35
Race/ethnicity			.07
White	18 (44)	15 (36)	
Black	17 (43)	13 (33)	
Hispanic	33 (82)	33 (81)	
Asian	28 (68)	29 (73)	
Other	4(10)	10 (25)	
Age			.13
18-34 y	16 (39)	12 (31)	
35-49 y	24 (61)	26 (66)	
50-64 y	31 (78)	39 (98)	
65-74 y	21 (52)	14 (34)	
75+ y	8 (21)	8 (21)	

Values are % (n).

^aDetails about distribution prior to dichotomizing results appear in [Supplementary Table 1](#).

with high uncertainty (Figure 3). In low-uncertainty cases, clinicians reported higher confidence in 58% (n = 111 of 190) of control cases vs 64% (n = 122 of 192) of intervention cases (OR, 1.33; 95% CI, 0.87-1.87; $P = .22$). In high-uncertainty cases, clinicians reported higher confidence in 9% (n = 6 of 65) of control cases vs 23% (14

of 62) of intervention cases (OR, 2.87; 95% CI, 1.02-8.13; $P = .048$). Similarly, adjusted analyses (Supplementary Table 4) found that the intervention impacted diagnostic confidence only in high-uncertainty cases (OR, 3.23; 95% CI, 1.09-9.52; $P = .03$) but not low-uncertainty cases (OR, 1.36; 95% CI, 0.86-2.16; $P = .19$).

Odds of intervention being associated with high confidence

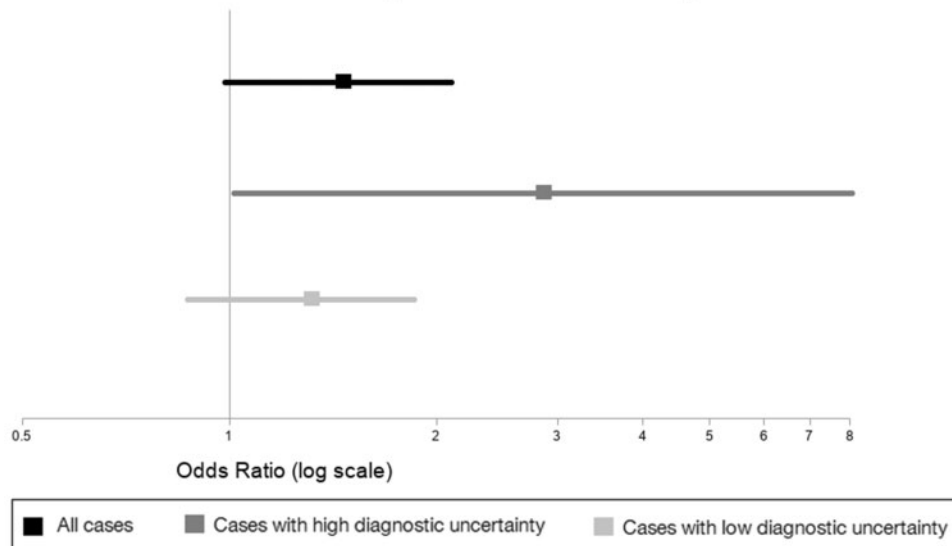


Figure 3. Unadjusted odds of peer input being associated with higher diagnostic confidence in all cases, cases with high diagnostic uncertainty, and cases with low diagnostic uncertainty.

Of the 255 intervention cases, clinicians reported that the collective opinion changed their plan in 45 (18%) cases and influenced their decision making in 84 (33%) cases. A higher rate of clinicians reported an impact on decision making in high-uncertainty ($n = 32$ of 62, 52%) vs low-uncertainty ($n = 52$ of 193, 27%) cases (OR, 3.57; 95% CI, 1.76-7.25; $P < .001$)

DISCUSSION

Key findings and implications

Clinicians report moderate or high diagnostic uncertainty in ~25% of outpatient encounters with potential for diagnostic uncertainty. Peer input did not increase diagnostic confidence in the primary unadjusted analyses. In exploratory analyses that adjusted for clinician characteristics and case perceptions, the peer input intervention was associated with higher odds of high confidence. Our subgroup analyses also demonstrated that the intervention was most impactful in cases when clinicians felt high baseline uncertainty, consistent with findings that clinicians most desire input on cases with continued uncertainty.¹⁹ We expand on prior literature by demonstrating that peer input on real ambulatory cases impacts diagnostic confidence and decision making in high-uncertainty cases.

Increasing diagnostic confidence has important implications.^{2,7,9} Uncertainty is associated with increased use of resources or tests.^{5,6} Thus, increasing confidence in high-uncertainty cases may reduce resource allocation. This is also validated in our prior work where clinicians reported that peer input consistent with their clinical decision making provides reassurance that extensive workups are not necessary.¹⁹

Peer input and collaboration are recommended approaches to improve diagnosis,³ and the pragmatic nature of this trial provides lessons on the feasibility of and barriers to digital collaboration. We believe that advances in informatics can address some of the barriers encountered in this study. First, case selection and input did not require PCPs to modify their normal workflows; scribes identified cases and relevant data based on chart review. This suggests that

with advances in interoperable data standards and natural language processing, peer input could be automatically solicited for cases with potential for diagnostic uncertainty. There were delays related to collecting enough responses to create the collective opinion; future collaboration interventions will need to identify facilitators of peer input to ensure real-time feedback from peers is more attainable. Also, since the digital platform allowed free-text data input, prior to returning the collective opinion to PCPs, clinicians at the digital platform manually matched terms with the same meaning that were not identified by an algorithm (eg, accidental acetaminophen poisoning and acetaminophen overdose). This demonstrates the importance of further development and dissemination of data standards to improve the utility of digital tools. Many delays were related to PCP actions (completing clinical notes; reviewing the input through an outside website); as technical solutions develop, such as automatic charting from conversations²² or electronic health record integration of external clinical decision support tools to allow access when needed (as recommended by experts),²³ delivery of real-time peer input will be increasingly attainable. Despite delays in the intervention, which meant that PCPs likely had additional information from the already initiated workup, we still found that peer input impacted confidence, suggesting that even if challenges to real-time input are not easily overcome, peer input still has value.

Limitations

This study is limited by our sample size though we recruited from 2 health systems and included 500+ encounters. We were underpowered for our primary analysis due to challenges with recruitment but still found in adjusted analysis that peer input impacted diagnostic confidence in high-uncertainty cases. We did not validate case information placed in the platform prior to soliciting input, and there were some deviations from our protocol in intervention timing and outcome assessment. Owing to data collection limitations, we were not able to document in detail the workflow steps that produced the greatest delays in intervention timing. However, this pragmatic design and its associated challenges increases generalizability about

how this could be implemented in clinical care.¹⁵ Given limited prior knowledge about diagnostic uncertainty in ambulatory care and our circumscribed follow-up time, we were not powered to study clinical outcomes.

CONCLUSION

This study contributes to literature by finding that in ambulatory encounters with new or unresolved complaints, diagnostic uncertainty is common. Further, we demonstrate that peer diagnostic input on ambulatory cases impacts diagnostic confidence in high-uncertainty encounters. Next steps include research that evaluates the impact of this intervention in other clinical settings, feasibility of implementing closer to real-time with fewer resources, and impact of input on clinical outcomes (eg, time to diagnosis, diagnostic accuracy), resource utilization, and clinician or patient-reported outcomes.

FUNDING

This work was supported by the Gordon and Betty Moore Foundation through a subaward from the Human Diagnosis Project (Grant No. 5496) (to US and VF), the National Heart Lung and Blood Institute of the National Institutes of Health (NIH) under Award Nos. K12HL138046 (to EK) and K23HL136899 (to VF), the National Center for Advancing Translational Sciences of the NIH under Award No. KL2TR001870 (to EK), and National Institute for Health's National Research Service Award (Grant No. T32HP19025) (to EK), and the National Cancer Institute (Grant No. K24CA212294) (to US). The content is solely the responsibility of the authors and do not necessarily represent the official views of the study funders or the NIH.

AUTHOR CONTRIBUTIONS

ECK, VF, SN, CRL, and US were involved in conception or design of the work. MH was involved in data collection. ECK, VF, NAR, CRL, and US were involved in data analysis and interpretation. ECK and VF were involved in drafting the article. All authors were involved in critical revision of the article and final approval of the version to be published.

SUPPLEMENTARY MATERIAL

[Supplementary material](#) is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

We would like to acknowledge our Human Diagnosis Project collaborators who provided us with the data for this project as well as the Human Diagnosis Project community for contributing responses. We also acknowledge Sarah Nouri, Kristan Olazo, and Kate Radcliffe for providing support for study procedures and data collection.

CONFLICT OF INTEREST STATEMENT

SN was an employee of the Human Diagnosis Project community during the study design phase of this project.

REFERENCES

- Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf* 2014; 23 (9): 727–31.
- Zwaan L, Hautz WE. Bridging the gap between uncertainty, confidence and diagnostic accuracy: calibration is key. *BMJ Qual Saf* 2019; 28 (5): 352–5.
- National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*. Washington (DC): National Academies Press; 2015.
- Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008; 121 (5): S2–23.
- Meyer AND, Payne VL, Meeks DW, et al. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med* 2013; 173 (21): 1952–61.
- Lave JR, Bankowitz RA, Hughes-Cromwick P, et al. Diagnostic certainty and hospital resource use. *Cost Qual Q J* 1997; 3: 26–32; quiz 46.
- Cifu AS. Diagnostic errors and diagnostic calibration. *JAMA* 2017; 318 (10): 905–6.
- Kostopoulou O, Rosen A, Round T, et al. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. *Br J Gen Pract* 2015; 65 (630): e49–54.
- Meyer AND, Singh H. Calibrating how doctors think and seek information to minimise errors in diagnosis. *BMJ Qual Saf* 2017; 26 (6): 436–8.
- Meyer AND, Singh H. The path to diagnostic excellence includes feedback to calibrate how clinicians think. *JAMA* 2019; 321 (8): 737–8.
- Nederhand ML, Tabbers HK, Splinter TAW, et al. The effect of performance standards and medical experience on diagnostic calibration accuracy. *Health Prof Educ* 2018; 4 (4): 300–7.
- Nagendran M, Chen Y, Gordon AC. Real time self-rating of decision certainty by clinicians: a systematic review. *Clin Med* 2019; 19 (5): 369–74.
- Willman AS. Use of Web 2.0 tools and social media for continuous professional development among primary healthcare practitioners within the Defence Primary Healthcare: a qualitative review. *BMJ Mil Health* 2020; 166 (4): 232–5.
- Boulos MNK, Maramba I, Wheeler S. Wikis blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. *BMC Med Educ* 2006; 6 (1): 41.
- Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015; 350 (1): h2147.
- Fontil V, Khoong EC, Hoskote M, et al. Evaluation of a health information technology-enabled collective intelligence platform to improve diagnosis in primary care and urgent care settings: protocol for a pragmatic randomized controlled trial. *JMIR Res Protoc* 2019; 8 (8): e13151.
- Bhise V, Rajan SS, Sittig DF, et al. Electronic health record reviews to measure diagnostic uncertainty in primary care. *J Eval Clin Pract* 2018; 24 (3): 545–51.
- Barnett ML, Boddupalli D, Nundy S, et al. Comparative accuracy of diagnosis by collective intelligence of multiple physicians vs individual physicians. *JAMA Netw Open* 2019; 2 (3): e190096.
- Fontil V, Radcliffe K, Lyson HC, et al. Testing and improving the acceptability of a web-based platform for collective intelligence to improve diagnostic accuracy in primary care clinics. *JAMIA Open* 2019; 2 (1): 40–8.
- Ng CS, Palmer CR. Analysis of diagnostic confidence: application to data from a prospective randomized controlled trial of CT for acute abdominal pain. *Acta Radiol* 2010; 51 (4): 368–74.
- Alam R, Cheraghi-Sohi S, Panagioti M, et al. Managing diagnostic uncertainty in primary care: a systematic critical review. *BMC Fam Pract* 2017; 18 (1): 79.
- Rajkomar A, Kannan A, Chen K, et al. Automatically charting symptoms from patient-physician conversations using machine learning. *JAMA Intern Med* 2019; 179 (6): 836.
- Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A Roadmap for National Action on Clinical Decision Support. *J Am Med Inform Assoc* 2007; 14 (2): 141–5.