JAMIA Open, 2(1), 2019, 40–48
doi: 10.1093/jamiaopen/ooy058
Advance Access Publication Date: 1 February 2019
Research and Applications

Research and Applications

Testing and improving the acceptability of a web-based platform for collective intelligence to improve diagnostic accuracy in primary care clinics

Valy Fontil,[1,2] Kate Radcliffe,[1,2] Helena C. Lyson,[1,2] Neda Ratanawongsa,[1,2] Courtney Lyles,[1,2] Delphine Tuot,[2,3] Kaeli Yuen[4] and Urmimala Sarkar[1,2]

[1]UCSF Division of General Internal Medicine, San Francisco, California, USA, [2]UCSF Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, San Francisco, California, USA, [3]UCSF Division of Nephrology, San Francisco, California, USA and [4]Keck School of Medicine, University of Southern California, Los Angeles, California, USA

Corresponding Author: Valy Fontil, MD, UCSF Division of General Internal Medicine, UCSF box 1364, San Francisco, CA 94143, USA (valy.fontil@ucsf.edu)

Received 21 August 2018; Revised 22 October 2018; Editorial Decision 21 November 2018; Accepted 5 December 2018

ABSTRACT

Objectives: Usable tools to support individual primary care clinicians in their diagnostic processes could help to reduce preventable harm from diagnostic errors. We conducted a formative study with primary care providers to identify key requisites to optimize the acceptability of 1 online collective intelligence platform (Human Diagnosis Project; Human Dx).

Materials and Methods: We conducted semistructured interviews with practicing primary care clinicians in a sample of the US community-based clinics to examine the acceptability and early usability of the collective intelligence online platform using standardized clinical cases and real-world clinical cases from the participants' own practice. We used an integrated inductive-deductive qualitative analysis approach to analyze the interview transcripts.

Results and Discussion: Perceived usefulness, perceived accuracy, quality assurance, trust, and ease of use emerged as essential domains of acceptability required for providers to use a collective intelligence tool in clinical practice. Participants conveyed that the collective opinion should: (1) contribute to their clinical reasoning, (2) boost their confidence, (3) be generated in a timely manner, and (4) be relevant to their clinical settings and use cases. Trust in the technology platform and the clinical accuracy of its collective intelligence output emerged as an incontrovertible requirement for user acceptance and engagement.

Conclusion: We documented key requisites to building a collective intelligence technology platform that is trustworthy, useful, and acceptable to target end users for assistance in the diagnostic process. These key lessons may be applicable to other provider-facing decision support platforms.

Key words: collective intelligence, human diagnosis project, diagnostic accuracy, diagnostic error, clinical reasoning

BACKGROUND AND SIGNIFICANCE

Diagnostic errors (defined as missed, delayed, or wrong diagnoses) in primary care impact an estimated 1 in 20 US adults every year.[1] These errors can lead to serious preventable harm, but few interventions have been developed and tested to reduce diagnostic errors in real-world primary care settings.[1–4] In current usual practice, providers commonly diagnose patients independently, without collaboration or consultation with other health professionals or use of health information technology (IT), leading to increased risk of diagnostic errors. In focus groups about outpatient diagnosis, physicians

identified use of technology to improve information availability and communication among providers as a key strategy to enhance timely and accurate diagnosis.[5] In its recent report on diagnostic error, the National Academy of Medicine suggested that health systems employ 2 key strategies essential to reducing diagnostic error in the ambulatory care setting: (1) enhance interprovider collaboration and (2) develop and utilize health IT innovations in the diagnosis process.[6]

Collective intelligence can be defined as shared or group intelligence that emerges from collaboration or collective efforts of many individuals. It harnesses the ability of a group to outperform the individual in a variety of cognitive tasks.[7] Indeed, obtaining a second opinion from a peer is a simple approach to improve clinical decision-making[8] IT platforms offer the opportunity to connect people and harness their collective intelligence. The use of collective intelligence among groups of primary care providers is a promising method to improve the accuracy of diagnoses and thereby reduce diagnostic error.[9,10] In discrete tasks related to medical decision-making, such as classification of radiology scans and pathological specimens, collective intelligence has been shown to improve accuracy when compared to individual decision-making.[11,12] However, improved diagnostic accuracy due to collective intelligence remains unproven in primary care practice. Studies examining technology-enabled collective intelligence platforms have shown that users are enthusiastic about crossf-discipline collaboration and easily obtaining expert feedback, but wary of inaccuracies and inefficiency of using a collective intelligence tool.[13]

One example of a collective intelligence platform for health-care professionals is the Human Diagnosis Project (Human Dx). In previous simulation testing, the collectively derived output from the Human Dx platform was shown to outperform its individual physicians in terms of suggesting the correct diagnosis in its assessment of clinical cases.[14] However, effective implementation of such a tool in primary care requires exploration of its usability and utility from the perspectives of practicing primary care clinicians. In this project, we conducted a formative study (ie, alpha testing) of the Human Dx collective intelligence tool among practicing primary care providers in a variety of community-based clinics in the United States to identify key requisites to optimizing the acceptability of the platform from the perspective of the potential end user—in this case, the primary care provider.

METHODS

Collective intelligence platform

The Human Diagnosis Project (Human Dx) is a web-based and mobile software designed to implement both key strategies for reducing diagnostic error recommended by the National Academy of Medicine—interprovider collaboration and use of health IT in the diagnosis process—by utilizing collective intelligence among physicians. Clinicians input the relevant details of a clinical case or question into the Human Dx platform. Then, any number of physicians (including peers and specialists) participating on the platform, typically 5 to 10, independently review the case and provide their own differential diagnoses and management plans. Human Dx aggregates the clinical assessments of the physicians and synthesizes them to produce a collective intelligence using advanced techniques including medical ontologies, semantic text extraction, and natural language processing.[14] The collective intelligence output from the tool, typically available within 24 h (Figure 1), is a summary with a list of diagnoses and suggestions for diagnostic workup.

Study design

We used semistructured interviews to examine acceptability and early usability of the collective intelligence online platform. This assessment was completed in 2 phases of testing; the first phase focused on standardized clinical cases and the second phase focused on real-world clinical cases from the participants' own practice (Figure 2). Platform enhancements to the Human Dx application were completed between the 2 phases in response to the first round of testing feedback. All providers were invited to complete both phases of testing, but were allowed to participate in a single round of testing based on their availability. The testing sessions lasted approximately 2 h per phase and were audio-recorded for transcription. Since the goal for this study was not to test the platform's performance but rather to determine key requisites for acceptability and usability, the information for both phases was collapsed together for analysis.

Phase 1: standardized cases

In the first phase, study participants independently evaluated 3 fictitious, standardized clinical case scenarios with varying level of diagnostic difficulty (see Supplementary Materials). Study participants provided their top 3 differential diagnoses, next steps, and level of confidence for each case. Then, a member of the study team entered each case into the Human Dx platform, which combined assessments from 5 to 10 Human Dx physician reviewers on the platform to generate the collective intelligence output that included a list of differential diagnoses and next steps. Physician reviewers were established users of the Human Dx platform, not recruited for the purpose of this study or hand-chosen by the study team. After reviewing outputs for all 3 cases, participants completed a semi-structured interview with study investigators informed in part by the technology acceptance model (TAM)[15] to assess broader opinions about the acceptability and usability of the platform. TAM is a validated theory of technology acceptance that has been widely used outside of health care and has become an important theoretical tool for health IT research.[15] As a theory, TAM suggests perceived usefulness and perceived ease of use as the 2 major factors that influence how users come to accept and use a technology. In addition to these concepts, our interview guide also included additional constructs such as general satisfaction and social influence (the perception of whether colleagues would accept or use collective intelligence) as well supplemental open questions to allow free expression of ideas. Participants were also asked to provide suggestions about the visual and content display of the platform output (see Supplementary Materials for contents of the interview guide).

Phase 2: participant-designated cases

In the second phase, participants independently evaluated 3 to 4 cases from their own practice for which they had a diagnostic question or ongoing diagnostic uncertainty. We did not provide any prompts or guidance for case selection. Participants who were outside San Francisco entered case details into the Human Dx platform after they created their own Human Dx accounts. San Francisco participants submitted a summary of case details to the study team to enter on their behalf into the Human Dx platform and the study team was approved to review the electronic health record if additional case details were needed. After the collective intelligence was generated for each case, participants reviewed the output generated by Human Dx, and participated in a semistructured interview with study investigators using the same interview guide as for standardized cases described above.

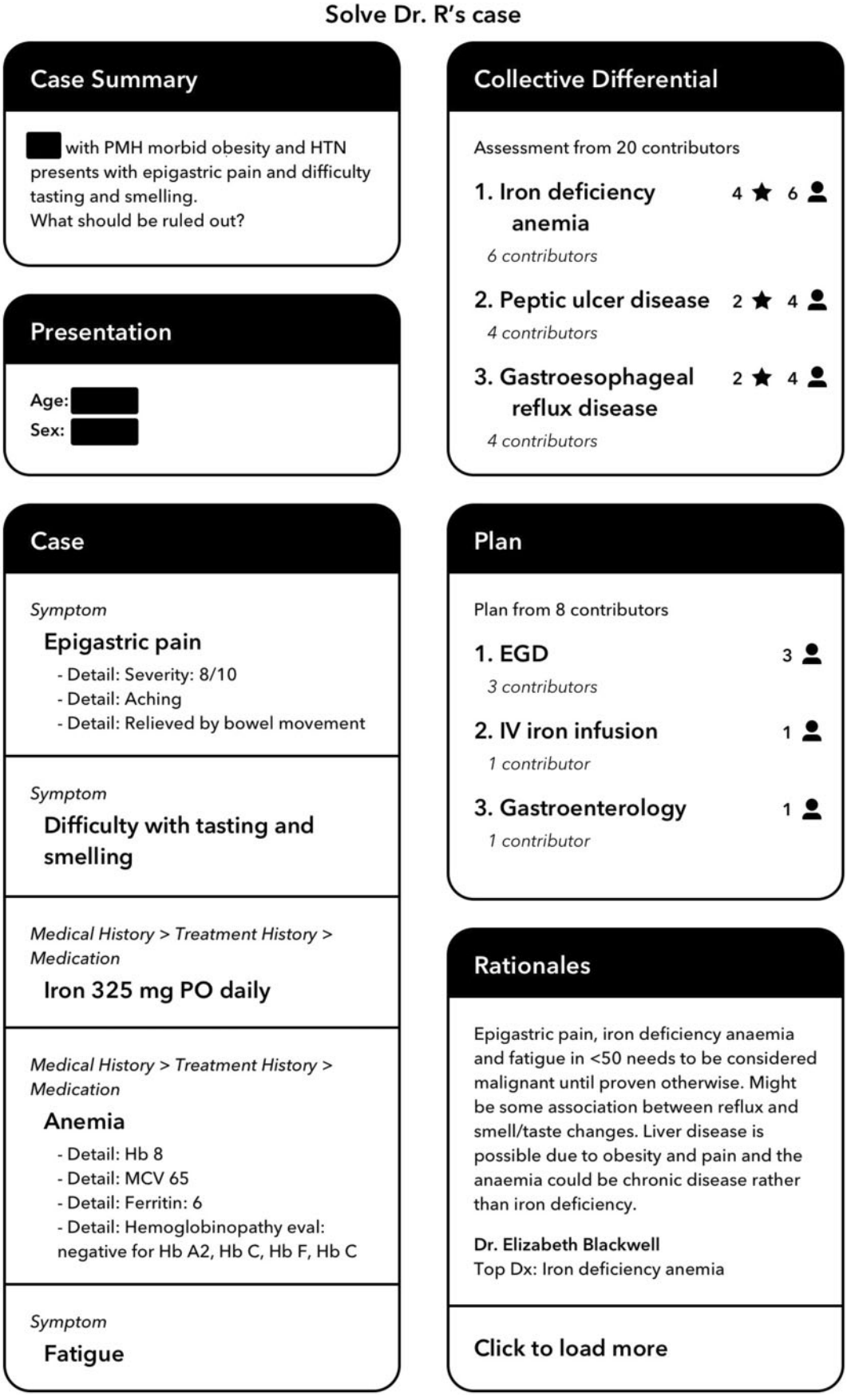

Figure 1. Screenshot of a collective intelligence output from the Human Dx platform for a clinical case. The left column displays the information entered by the participating clinician. The right column displays collective intelligence output with the differential diagnosis, plan, and rationales.

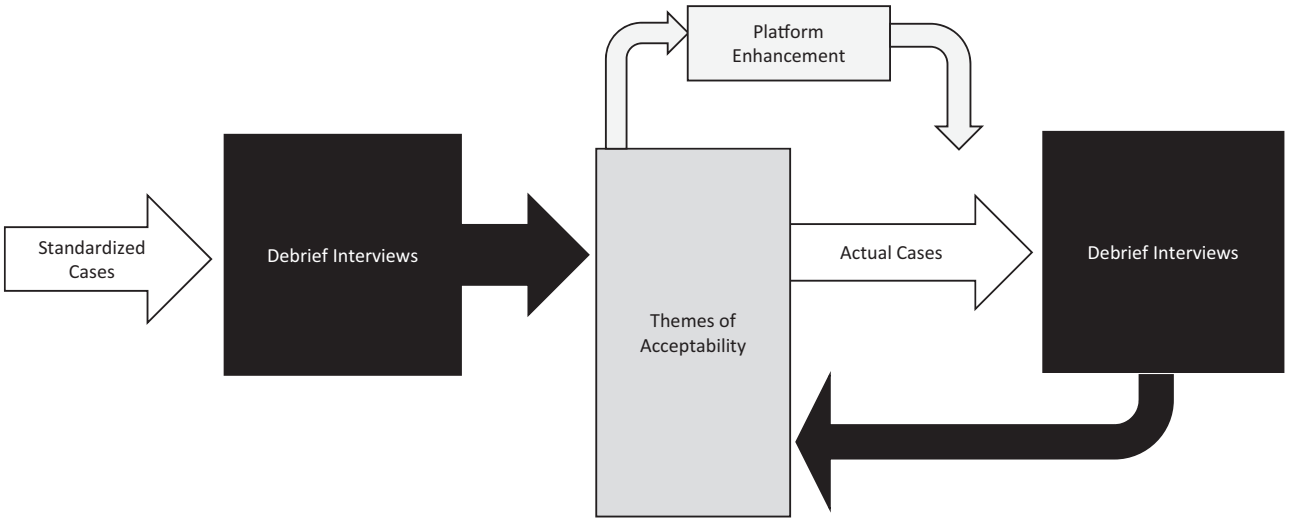

Figure 2. We used semi-structured debrief interviews to conduct acceptability and early usability testing of the collective intelligence online platform. This assessment was completed in two phases of testing using standardized clinical cases (phase 1) and real-world clinical cases from the participant's own practice (phase 2).

Study setting

We used a purposive sampling approach to recruit a diverse mix of primary care providers with respect to geographic location (urban, suburban, and rural), practice type (group and solo), and institutions (academic and nonteaching). This sampling strategy allowed for diversity in provider access to clinical specialists and other diagnostic testing services, to explore differences in perspectives from providers with more or less access to resources that support diagnosis. Eligible providers included physicians and nurse practitioners within our network of 12 primary care safety-net clinics in San Francisco and providers known to those within the network through snowball sampling wherein initial participants referred additional colleagues who practiced in clinical settings that differed from theirs (eg, Indian Health Service, solo suburban practice, and Saipan).

Data analysis

Qualitative analysis of the interview transcripts explored the acceptability of collective intelligence, with a focus on key requisite to achieve acceptability of routine and sustained use of the platform by providers in primary care clinical settings. Transcripts were coded using an integrated inductive-deductive qualitative data analysis approach.[16] In particular, we used the constant comparison method, an inductive qualitative data analysis approach in which data are broken down, compared for similarities and differences, and grouped together under similar conceptual themes[17] to uncover a wide variety of themes from the data, while also employing predetermined conceptual codes drawn from the TAM[18] that focuses on the perceived usefulness and ease of use of a new technology to structure a deductive analysis of the data. One study author (H.C.L.) independently coded the transcripts to identify preliminary themes through initial readings of the transcripts. Two additional coauthors (V.F. and K.R.) read all the transcripts independently and reviewed and made revisions to the initial coding. Iterative discussions with the full team of study investigators refined the thematic categories and coding that led to final set of salient themes and coding by consensus.

This study was exempted by the Institutional Review Board at the University of California, San Francisco as a formative phase in a quality improvement project.

RESULTS

We recruited 13 primary care providers (8 in San Francisco, 1 in Georgia, 1 in New Mexico, 1 in San Diego, and 2 in Saipan, a US commonwealth in the Northern Mariana Islands). Eight completed both phases, 3 completed only the first session with standardized cases, and 2 completed only the second phase with cases from their own practice.

Participating providers work in clinics that differ in access to diagnostic support resources, including access to specialist consultation via electronic consultation. All 8 participants located in San Francisco work in urban safety-net clinics that care for underserved populations and have access to electronic specialty consultation through their health system. The single participant in Georgia works in a solo practice in a medium-sized city, with no access to electronic specialty consultation. In New Mexico, the participant works for an Indian Health Service with access to in-person specialty support. The 2 participants in Saipan practice in a rural health clinic associated with a community hospital, with access to telemedicine for certain subspecialties and the ability to refer patients to the nearby island of Guam for in-person consultation. The San Diego participant practices in a suburban group practice that serves primarily working poor and elderly, with access to consultation. Two participants are Nurse Practitioners and the remaining 11 providers are physicians trained in family medicine or internal medicine (Table 1).

In participant interviews, several domains of acceptability emerged as essential elements required for providers to use a collective intelligence tool in clinical practice: perceived usefulness, perceived accuracy, quality assurance, trust, and ease of use. Our inductive qualitative analysis allowed us to arrive at refined definitions for these domains from the perspective of prospective end users—presented in Table 2. Within each domain, we identified several themes as key requisites in optimizing the acceptability of the collective opinion tool among primary care providers (Table 3).

Perceived usefulness

Participants raised several requisites to optimizing usefulness. They conveyed that the collective opinion should: (1) contribute to their clinical reasoning, (2) boost their confidence, (3) be generated in a

Table 1. Characteristics and practice settings of participating clinicians

Description	No. of participants ($n = 13$)
Participant training	
Internal medicine	8
Family medicine	3
Nurse practitioner	1
Physician assistant	1
Participant location	
San Francisco	8
Saipan	2
Georgia	1
New Mexico	1
San Diego	1
Years in practice	
Less than 5 years	6
5–10 years	2
10–20 years	1
More than 20 years	2
Clinic setting	
Urban health clinic	8
Rural health clinic	3
Medium-sized city solo practice	1
Medium-sized city group practice	1
Associated with a university	2
Associated with a hospital	3
Safety-net clinic	12
Access to specialty consultation	
Electronic consultation system embedded in Electronic Medical Record	9
Telemedicine for certain subspecialists, refer to the nearby island of Guam for in-person consultation	2
Telemedicine and in-person specialty access	1
No access to electronic specialty consultation	1

Table 2. Definitions for key domains of acceptability for technology-enabled collective intelligence

Domains of acceptability	Definitions
1. Perceived usefulness	The degree to which participants felt that collective intelligence added value to their work as a provider, or was helpful in diagnostic thinking or decision-making for a particular case
2. Perceived accuracy	The degree to which users found the list of diagnoses and recommendations provided in the collective intelligence output reasonable, accurate, and safe
3. Transparent quality assurance	The degree to which the technology platform provides information on the qualifications and expertise of its collective intelligence contributors and expertise is relevant to a case
4. Trust	End-user belief and confidence that the technology platform is legitimate, reliable, and able to consistently provide high-quality, accurate output to help clinical decision-making
5. Ease of use	The facility with which the user can enter cases into the platform and the anticipated efficiency of incorporating the process of case entry within the user's routine workflow

Note: We derived these definitions in part from qualitative analysis of the semistructured interviews with study participants.

timely manner, and (4) be relevant to their clinical settings and use-cases (Table 4).

Providers expect concrete cognitive contribution to clinical reasoning and decision-making as a utility for a collective opinion tool. The collective opinion should not just provide a list of possible answers to their diagnostic question. Details about the rationale for the answers given would be a significant value-add to their thinking process and decision-making.

> I think that if it can be honed and improved, especially in terms of. . .providing more clinical decision-making support and more information about why certain providers made that certain decision, I think then I would say it would probably be moderately to very helpful.

Affirmation of users' current diagnostic thinking is a valuable contribution to boosting confidence in decision-making. Study participants expressed finding significant value in a tool that bolsters their confidence and further expands their thinking process. When providers receive collective intelligence output that affirms their current thinking, it can influence their decision-making in terms of selecting appropriate diagnostic testing. In some instances, such affirmation can help providers avoid unnecessary testing and the financial costs associated with such testing.

> So, getting all of that [collective intelligence output] reassures me that she maybe doesn't need the $300 000 workup. So, seeing that [collective intelligence output] would make me more confident in addressing the anxiety right away before doing the million-dollar workup.

Receiving the collective intelligence in a timely manner is key to usefulness. In order for the collective intelligence to be useful to participants, it should be generated within a time period that allows participants to act on the recommendations. A delay could make it inconvenient, irrelevant, or inconsequential as they may use other resources for help or have moved their attention to other important clinical tasks.

> Then in terms of waiting 24 hours like for the collective opinion—A lot of my patients are driving several hours to come to each appointment or hitchhiking or it's a lot for them to come to each appointment. A lot of times I have to make decisions that day. So I think that would be tough.

End user clinical settings and guidance on ideal use-cases may be important to consider.

Participants expressed varied sentiments regarding use-cases and clinical settings for which a collective opinion would be most useful. Participants in an urban academic setting commented on the available resources that they would use instead of entering cases in the collective intelligence platform, or that the platform may be more useful in a case where they have already exhausted the available resources. On the other hand, rural providers were more likely to find the output more useful than their current resources (Table 4). In addition, participants varied in what types of cases they felt were appropriate. Some felt that only a small proportion of cases would benefit from the collective input.

> If I have a curveball case, it would be great to have the diagnosis. But oftentimes. . .Thinking back on the cases that I saw this week, how many would I use the [tool] on?

Table 3. Summarized keys to acceptability and related potential pitfalls to avoid based on analysis of interviews

Keys to acceptability	Potential pitfalls
Trust in quality of contributors and accuracy of the collective intelligence output	Avoid unreasonable, inappropriate, or irrelevant recommendations Avoid contributors that can be perceived as unqualified (overall or for a specific case)
Importance of cognitive contribution to provider clinical thinking or decision-making	Avoid output that fails to enhance users' thinking process or help with next steps in diagnostic decision-making
Importance of timeliness of content	Delayed feedback may be difficult to incorporate in usual workflow for diagnostic decisions
Education on best use cases	Insufficient guidance or training on appropriate target use-case scenarios can lead to infrequent or inappropriate use of a collective intelligence technology platform
Ease of use	Avoid cumbersome and time-intensive user procedures

Note: This table is a summarized interpretation drawing key requisites for development of clinician-facing collective intelligence technology platforms based on the themes from our qualitative analysis outlined in Table 2.

Trust

Consistent accuracy and quality assurance are essential to engender and maintain end-user trust.

The issue of trust emerged as one of the most important factors to whether participants would likely use a collective intelligence platform in their practice (Table 4). Participants pointed out 2 factors as key to earning and maintaining their trust: accuracy and quality assurance. Accuracy of the platform is a critical component to participants' trust in the collective intelligence, affecting whether they intend to use the tool. Any seemingly unreasonable, irrelevant, or inappropriate suggestion in the collective opinion would erode participants' trust in the platform and their likelihood to use it in the future.

> Regardless of everything, it has to be accurate. If it's not accurate and it's not fast, then I'd have resources that already exist that are much better.

With respect to quality assurance, participants want to know more information about the community of physician contributors on the platform. In the course of the interviews, study participants asked about Human Dx contributors' qualifications, how they are recruited, how they are compensated, and how they are evaluated. They also want more information on how Human Dx uses artificial intelligence to compile and rank diagnoses into the collective intelligence list. Without this information, participants expressed that they could not trust the collective intelligence.

> I don't trust [the tool] because I don't know who [the contributors] are. I don't know what their training is, I don't know what evidence they're using to make these decisions.

Ease of use

While the interviews primarily discussed review of the output, a few participants entered their own cases into the platform and concerns about the process to enter cases into the platform came up in the discussion (Table 4).

Inputting data into the platform should be straightforward and output should be understandable. In order for participants to use a collective intelligence tool, they expressed that the tool should be simple and efficient.

> The tool itself could be a little easier to use and a little differently organized... Ease of input is a huge factor. If it's going to take me 10, 15 minutes to put the information in, I'll most likely not use it.

In reviewing the collective intelligence output, participants similarly expressed that the tool should be clear and easy to navigate.

DISCUSSION

We sought practicing clinician perspectives on using a collective intelligence technology tool that synthesizes input from other clinicians, theoretically peers, to assist with the cognitive process of diagnosis. In that sense, this platform encompasses both a peer network and a technology. We found that trust in the peer network and the clinical accuracy of the collective intelligence output were prerequisites for engaging with the technology. Other keys to acceptability included timeliness of the output, the importance of actionable recommendations that contribute to users' clinical thinking and decision-making, education/training on appropriate use-cases, and efficient procedures that fit within usual clinical workflow. Participants also suggested use of collective intelligence technology could be particularly valuable in clinical settings with limited access to peer or specialty consultation.

At present, collective intelligence is rarely used in medical decision-making,[1,11,19–21] and to our knowledge, this study is the first to test the potential of such an approach to assist primary care providers in their cognitive process of making a diagnosis.

This study may also be the first to employ the TAM to assess end-users acceptance or willingness to use collective intelligence in clinical practice and makes an important addition to the body of literature in this regard. TAM has been used to assess physicians' acceptance of various health IT platforms including telemedicine, communications systems, computerized provider order entry, mobile health technologies, and computerized clinical decision support tools.[15] While these previous studies used TAM to evaluate user acceptance of existing health IT, our study is also unique in using TAM early in the technology development process to uncover key requisites for developing a clinician-facing health IT tool for assistance in the cognitive process of diagnostic assessment and decision-making.

Our findings suggest that the TAM could be extended by explicitly specifying trust as a potential contributing factor for "perceived usefulness" and core component of acceptability for technology solutions that rely on collective intelligence and peer-based feedback. Although we did not set out to develop or propose a novel conceptual framework for technology acceptance, our findings suggest that a modification of TAM to include trust (predicated on perceived accuracy and quality assurance) could be warranted to serve as a guide for deployment of collective opinion tools and perhaps other

Table 4. Primary care providers' reactions regarding the acceptability and potential usability of technology-enabled collective opinion

Domains of acceptability	Themes	Quotes
1. Perceived usefulness	Providers expect concrete cognitive contribution to clinical reasoning and decision-making as a utility.	"I think that if it can be honed and improved, especially in terms of providing multiple next steps or even providing more clinical decision-making support and more information about why certain providers made that certain decision, I think then I would say it would probably be moderately to very helpful."
		Where this tool would be really helpful is if somebody's able to display how they solved the problem. . .I would just make it so you can click on the diagnosis and read the opinions of the five people with the evidence."
		"The way that this helps me is that it makes me look into areas that I previously may not have looked into."
		"I think that's why I would consider using this tool just because it can help me broaden my thinking process about things or even trigger new ideas about the case. So that is what I'm looking for."
	Affirmation of users' current diagnostic thinking is a valuable contribution to boosting confidence in decision-making.	"Just having the differential diagnosis there all within the lines that I was thinking about. It influenced me keeping with the work-up. . .to order the test that I was thinking about. It definitely influenced it knowing that other people were thinking of my same diagnoses before I get the work-up."
		"So, getting all of that [output] reassures me that she maybe doesn't need the $300 000 workup and the anxiety is playing a big role in this. When I started to see the results coming in, I was like, 'Oh, I wonder,' because she does have findings but there's also this element of anxiety. So, seeing that [output] would make me more confident in addressing the anxiety right away before doing the million-dollar workup."
		"I think in some ways I felt satisfied that it seemed like there were some agreements with what I was thinking."
	Receiving the collective intelligence in a timely manner is key to usefulness.	"More rapid turnaround time would make [the tool] a lot more useful. . .I mean, before the patient leaves the office."
		"I think if it was like instantaneous or like in 30 minutes or something then I think it would be much more useful."
		"Ideally if [the results] were within the same business day or within the same half day would be nice."
	Ideal use-cases and clinical settings may be important to consider.	"I think [using the tool] is going to be [for] the cases where I'm just kind of stumped and I don't know who to ask, something really complex that doesn't have a well-defined problem for one specialist."
		"I think it adds value specifically to cases where you think that you've done all the necessary work up and you're still not entirely sure. Especially in this rural area that I'm working in, where I don't quite have a lot of consultants to ask or a lot of available colleagues to discuss cases about, that actual reassurance is valuable."
		I think without having somebody to bounce ideas off of, it is really nice just to have reassurance that I'm on the right track even for that. It's useful."
2. Perceived accuracy	Providing output that is consistently reasonable, accurate, and safe is paramount to establish and maintain trust.	"Regardless of everything, it has to be accurate. If it's not accurate and it's not fast, then I'd have resources that already exist that are much better."
		"I think that if I trusted that the information I'm getting is consistently accurate and seems to- even ring true based on the cases, I think that then I would feel more confident in using it."
		"I'm beginning to lose confidence. . .because I don't really know where you get—[the output] just didn't make any sense at all"
3. Transparent quality assurance	Uncertainty about the competence or relevant expertise of contributors on the platform can erode trust.	"I worry about who these people are. . .are these just volunteers? Are they paid? Who's sitting there and doing this?"
		"It could be nice to see if somebody is always suggesting the right thing and then the accuracy of that. . .If you had a way of designating one of those contributors who was getting the correct diagnoses over 80% of the time. I think that would be helpful."
		"I think it would be a little bit more settling to know that if there's somebody who is a specialist in a certain area who is eyeballing certain cases. I'm just not very convinced by seeing a dermatologist answer a question about chest pain. That really doesn't help persuade me."
		"I like that [the tool] tells you whether [the consultants] are like PCPs or dermatologists or like the surgeon. I think that breakdown is helpful just cause like for me I'm coming from a PCP background so I maybe would, depending on the case, out more weight on who's suggesting what."

(continued)

Table 4. continued

Domains of acceptability	Themes	Quotes
4. Trust	Consistent accuracy and quality assurance are essential to engender and maintain end-user trust.	"I did wonder. . .I don't know who these PCPs are. I don't know if they're well-qualified. I don't know if I trust their opinions. So I do think it's possible to make the wrong decisions based on this app." "I don't trust [the tool] because I don't know who [the contributors] are. I don't know what their training is, I don't know what evidence they're using to make these decisions." "I think that if I trusted that the information I'm getting is consistently accurate and seems to- even ring true based on the cases, I think that then I would feel more confident in using it."
5. Ease of use	Inputting data into platform should be straightforward and output should be understandable.	"I don't know how you input all these data but it is—it looks a little bit maybe tedious. If you're checking through chest pain, how long etc. and you've already typed that into your note, you're not going to want to type it in somewhere else or enter it somewhere else so I can see that being a barrier." "The tool itself could be a little easier to use and a little differently organized. . . Ease of input is a huge factor. If it's going to take me 10, 15 minutes to put the information in, I'll most likely not use it."

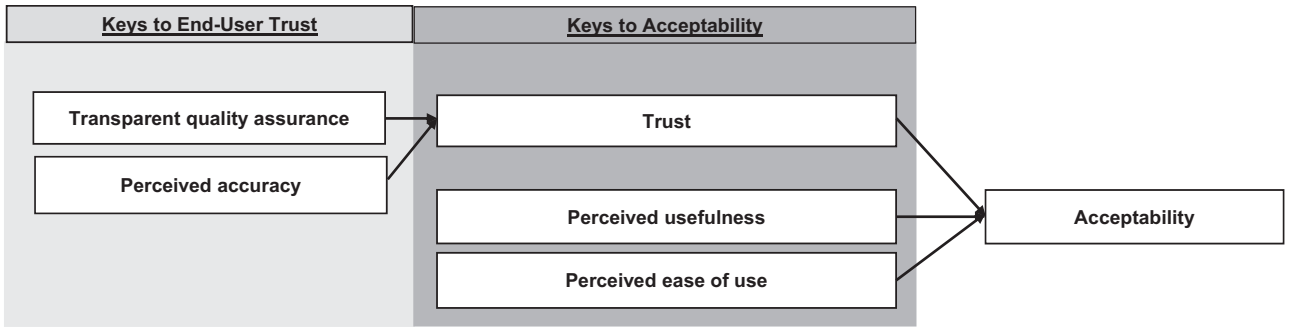

Figure 3. Proposed modified Technology Acceptance Model for collective intelligence technology with trust added as potential contributing factor to perceived usefulness and a key factor for acceptability.

provider-facing technologies such as computerized clinical decision support tools (see Figure 3). For these types of technological interventions, trust appears to be an incontrovertible prerequisite for acceptance. In other words, if clinicians are not assured that the tool is medically and scientifically accurate and that quality measures are in place to ensure accuracy, they will not likely use it.

There are some limitations to consider in interpreting our findings. We designed this study as an acceptability study in a limited sample of primary care providers. Hence, our findings may not be generalizable to all providers or clinical settings. Potential differences by provider characteristic (eg, type of professional degree, years of experience, specialty, etc.) could not be examined due to the limited sample size of participants.

CONCLUSION

We documented key requisites to designing a collective intelligence technology platform that is trustworthy, useful, and acceptable to target end users—primary care providers—for assistance in the diagnostic process. These key lessons will contribute to iterative development and testing of the platform in a future demonstration study and may be applicable to other provider-facing decision support platforms. Optimization of a collective intelligence tool has the potential to help primary care providers make more timely and accurate diagnoses by increasing access to peer and specialty consultation, particularly in clinical settings that have limited resources.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONTRIBUTORSHIP STATEMENT

All authors have fulfilled the criteria for authorship established by the International Committee of Medical Journal Editors and approved submission of the manuscript. V.F. and U.S. made substantial contributions to the conception and design of the study and drafting of the manuscript. C.L. contributed substantially to the study design and qualitative analysis. H.C.L. conducted the initial qualitative coding of interview transcripts. K.R. and V.F. designed and drafted interview guides with guidance from C.L. and conducted the semistructured interviews with study participants. K.Y.

contributed to data acquisition by organizing the content and user-interface design of the output generated by the Human Dx platform. D.T. and N.R. made important intellectual contributions to study design as experts in electronic consultation platforms and health information technology. V.F., K.R., H.L., N.R., C.L., D.T., and U.S. made substantial contributions to iterative qualitative data analysis and interpretation. All coauthors participated in revising the manuscript critically, made important intellectual contributions, and approved the final version to be published.

ACKNOWLEDGMENT

The authors acknowledge the important contributions of Mekhala Hoskote in her administrative role in conducting this research.

FUNDING

This work was supported by the Gordon and Betty Moore Foundation through a subaward from the Human Diagnosis Project, K23HL136899, AHRQ Grants 1K08HS022561 and P30HS023558.

Conflict of interest statement. K.Y., who worked on this project as a medical student, has become employed by the Human Diagnosis Project after completing her role in data collection and coauthor for this work. The other authors have no competing interest to report. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Singh H, Meyer AN, 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.
2. Riches N, Panagioti M, Alam R, *et al*. The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis. *PLoS One* 2016; 11 (3): e0148991.
3. Graber ML, Kissam S, Payne VL, *et al*. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012; 21 (7): 535–57.
4. Nurek M, Kostopoulou O, Delaney BC, Esmail A. Reducing diagnostic errors in primary care. A systematic meta-review of computerized diagnostic decision support systems by the LINNEAUS collaboration on patient safety in primary care. *Eur J Gen Pract* 2015; 21 Suppl: 8–13.
5. Sarkar U, Simchowitz B, Bonacum D, *et al*. A qualitative analysis of physician perspectives on missed and delayed outpatient diagnosis: the focus on system-related factors. *Jt Comm J Qual Patient Saf* 2014; 40 (10): 461.
6. Ball JR, Balogh E. Improving diagnosis in health care: highlights of a report from the national academies of sciences, engineering, and medicine. *Ann Intern Med* 2016; 164 (1): 59–61.
7. Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW. Evidence for a collective intelligence factor in the performance of human groups. *Science* 2010; 330 (6004): 686–88.
8. Payne VL, Singh H, Meyer AN, Levy L, Harrison D, Graber ML. Patient-initiated second opinions: systematic review of characteristics and impact on diagnosis, treatment, and satisfaction. *Mayo Clin Proc* 2014; 89 (5): 687–96.
9. Kammer JE, Hautz WE, Herzog SM, Kunina-Habenicht O, Kurvers R. The potential of collective intelligence in emergency medicine: pooling medical students' independent decisions improves diagnostic performance. *Med Decis Making* 2017; 37 (6): 715–24.
10. Hautz WE, Kammer JE, Schauber SK, Spies CD, Gaissmaier W. Diagnostic performance by medical students working individually or in teams. *JAMA* 2015; 313 (3): 303–4.
11. Kurvers RH, Krause J, Argenziano G, Zalaudek I, Wolf M. Detection accuracy of collective intelligence assessments for skin cancer diagnosis. *JAMA Dermatol* 2015; 151 (12): 1346–53.
12. Wolf M, Krause J, Carney PA, Bogart A, Kurvers RH. Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. *PLoS One* 2015; 10 (8): e0134269.
13. Sims MH, Bigham J, Kautz H, Halterman MW. Crowdsourcing medical expertise in near real time. *J Hosp Med* 2014; 9 (7): 451–6.
14. Abbasi J. Shantanu Nundy, MD: the human diagnosis project. *JAMA* 2018; 319 (4): 329–31.
15. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. *J Biomed Inform* 2010; 43 (1): 159–72.
16. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv Res* 2007; 42 (4): 1758–72.
17. Corbin J, Strauss A. *Basics of Qualitative Research*. Los Angeles, CA: Sage; 2014.
18. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. *MIS Quart* 2003; 27 (3): 425–78.
19. Hukkinen K, Kivisaari L, Vehmas T. Impact of the number of readers on mammography interpretation. *Acta Radiol* 2006; 47 (7): 655–9.
20. Duijm LE, Louwman MW, Groenewoud JH, van de Poll-Franse LV, Fracheboud J, Coebergh JW. Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. *Br J Cancer* 2009; 100 (6): 901–7.
21. Farnetani F, Scope A, Braun RP, *et al*. Skin cancer diagnosis with reflectance confocal microscopy: reproducibility of feature recognition and accuracy of diagnosis. *JAMA Dermatol* 2015; 151 (10): 1075–80.