

## Patient information needs for transparent and trustworthy artificial intelligence in healthcare

Austin M. Stroud, MA<sup>1</sup>, Sarah A. Minter, PhD<sup>2</sup>, Xuan Zhu, PhD, MS<sup>3</sup>, Jennifer L. Ridgeway, PhD<sup>3, 4</sup>,  
Jennifer E. Miller, PhD<sup>5</sup>, Barbara A. Barry, PhD<sup>3, 4</sup>

<sup>1</sup>Biomedical Ethics Program, Mayo Clinic, Rochester, MN, USA 55905

<sup>2</sup>Physical Medicine and Rehabilitation Research, Mayo Clinic, Rochester, MN, USA 55905

<sup>3</sup>Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA 55905

<sup>4</sup>Division of Health Care Delivery Research, Mayo Clinic, Rochester MN, USA 55905

<sup>5</sup>Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven CT, USA 06520

### Correspondence:

Austin M. Stroud  
200 First Street SW  
Rochester, MN 55905  
Email: [stroud.austin@mayo.edu](mailto:stroud.austin@mayo.edu)

## Abstract

**Background:** As health systems incorporate artificial intelligence (AI) into various aspects of patient care, there is growing interest in understanding how to ensure transparent and trustworthy implementation. However, little attention has been given to what information patients need about these technologies to promote transparency of their use.

**Methods:** We conducted three asynchronous online focus groups with 42 patients across the United States discussing perspectives on their information needs for trust and uptake of AI, focusing on its use in cardiovascular care. Data were analyzed using a rapid content analysis approach.

**Results:** Our results suggest that patients have a set of core information needs, including specific information factors pertaining to the AI model, oversight, and healthcare experience, that are relevant to calibrating trust as well as perspectives concerning information delivery, disclosure, consent, and physician AI use.

**Conclusions:** Identifying patient information needs is a critical starting point for calibrating trust in healthcare AI systems and designing strategies for information delivery. These findings highlight the importance of patient-centered engagement when considering approaches for transparent healthcare AI.

**Keywords:** Artificial intelligence, Machine learning, Health communications, Qualitative, Cardiovascular

## Introduction

A chief concern among published ethical frameworks and guidance documents for responsible healthcare artificial intelligence (AI) is the lack of transparency regarding how AI models—computer software trained to recognize patterns in health data—have been developed and trained.<sup>1-7</sup> These models are often opaque without insight into how or why a particular output was derived, raising concerns about transparency even if the output proves to be measurably accurate.<sup>8</sup> This “black-box” problem poses a unique challenge in healthcare settings, where well-informed and transparent clinical decision-making is critical to healthcare delivery and patient trust.<sup>9</sup> There are also implications for health equity, with greater transparency being a potential mechanism for mitigating bias in models.<sup>10</sup> Furthermore, physicians may need to grapple with appropriate disclosure of AI use and associated risks, particularly when details about model development and training are lacking; nonetheless, concerns exist that insufficient AI transparency may be disruptive to physician-patient relationships which are essential for determining proper clinical investigation and treatment in response to AI outputs.<sup>11-13</sup>

Efforts to facilitate transparency for healthcare AI tools include novel technical approaches to illuminating model development processes and creating more explainable AI (XAI) tools.<sup>14, 15</sup> For example, visualizations for deep learning systems in cardiovascular disease modeling aim to support understanding by highlighting interactions between variables.<sup>16</sup> However, these technical approaches are limited and XAI only serves as one mode for enhancing transparency in a system with numerous stakeholders of varying information needs, particularly patients who may lack the technical skills and training to interpret these outputs.<sup>17</sup>

In addition to technical approaches, there is growing scholarship around AI model documentation—the provision of information accompanying AI tools for enhanced transparency.<sup>18, 19</sup> These approaches provide supplemental information and evaluative criteria about a given model or data set and, in some instances, can be dynamically tailored for different use cases or revised as models are updated. Prototypes such as “AI nutrition labels” that display information similar to nutrition or prescription drug labels have been developed for technical and clinical end-users (i.e., software engineers, data scientists, and physicians). However, there is still a lack of consensus regarding the optimal approach for including and assessing information about healthcare AI, and these efforts have largely focused on enhancing transparency for a technical and clinical expert audience.<sup>20-22</sup>

There is limited empirical work focusing on enhancing transparency of AI in healthcare for patients. Recent qualitative research exploring AI information has examined the needs of patients regarding diabetes treatment, including creating tutorials for information access and alerts to guide device use.<sup>23</sup> Other research assessing patient perspectives on AI in healthcare more broadly, or for specific tools, has identified patient concerns about transparency, including the potential for models to make unexplainable errors or limit patient decision-making.<sup>24-26</sup> While current scholarship emphasizes transparency as a key principle, there is limited evidence on patients’ specific information needs. This knowledge gap creates a significant barrier to developing inclusive and transparent solutions and enhancing understanding of information factors—discrete elements of information communicated to a patient (e.g., model performance, manufacturer)—that influence patient trust in healthcare AI. There are also downstream implications for safety and efficacy, as model transparency may factor into clinician and patient decisions about AI use and related clinical recommendations.<sup>27</sup>

The present online focus group study aimed to identify patients’ core information needs if AI tools were to be used in their health care. To elicit patient perspectives, we used clinical vignettes depicting AI in the diagnosis, treatment, and/or monitoring of cardiovascular conditions and posed questions about healthcare AI in general. We also explored information factors that influence patient trust in these

technologies, their preferences for information delivery, and their perspectives about physicians using AI in their care. Our overall objective is to provide a more robust evidentiary foundation for designing patient-centered transparency solutions for healthcare AI and calibrating trust in these tools.

## Methods

### Study Design

We conducted asynchronous online focus groups with patients across the United States to identify their information needs and preferences to consider transparent and trustworthy AI. Clinical vignettes depicting AI technologies in cardiovascular care were used as prompts for discussions along with general questions about healthcare AI broadly. This study is part of a larger multi-method investigation of clinician and patient information needs and ethical considerations for adoption of AI in healthcare.

### Ethics

Review and approval for this study and all procedures was obtained from the Mayo Clinic Institutional Review Board (protocol # 21-012302). Consent was obtained from participants verbally prior to each focus group in accordance with IRB guidance.

### Recruitment

Potential research participants were sampled via ResearchMatch, a national health volunteer registry with over 140,000 subjects who have agreed to be contacted about research opportunities. Several academic institutions created this database; it is supported by the U.S. National Institutes of Health as part of the Clinical Translational Science Award (CTSA) program.

We contacted participants via the ResearchMatch platform with a message containing a brief description of our study, eligibility criteria, and voting buttons that let subjects affirm or deny their interest. ResearchMatch provided contact information (e.g., name, email address) for those who responded affirmatively to the message. These eligible participants were added to our recruitment list. Participants were considered eligible if they were at least 18 years of age, English-literate, and had a primary care or cardiology clinical visit within the past three years. We oversampled from American Indian, Asian, Black/African-American, Native Hawaiian/Pacific Islander, Hispanic/Latino, and Multi-racial populations to increase their representation in our sample, as these groups have been underrepresented in US medical research.<sup>28, 29</sup>

### Data Collection

Three asynchronous online focus groups were conducted.<sup>30</sup> Each focus group took place over two days on FocusGroupIt (focusgroupit.com). Focus groups were limited in size to up to 15 participants based on best practices that balances the depth and richness of discussion and participant and moderator burden.<sup>30, 31</sup> Questions were pre-loaded into the FocusGroupIt platform by a member of the research team (AMS). The first set of questions were posted at midnight, and the second set of questions were posted 24 hours later. Participants responded to both free-response and multiple-choice questions to enhance the breadth of collected data. Some questions were presented alongside a hypothetical clinical vignette describing AI used in the care of a cardiovascular condition; a vignette example is depicted in Box 1. These vignettes were reviewed by subject matter experts for clinical accuracy and patient-centered communication.<sup>32</sup> Participants reviewed four total vignettes adapted from cardiovascular technologies in industry and academic literature. Technologies depicted included 1) a smartwatch that monitored ECG readings, 2) a stethoscope that classified heart murmurs, 3) a monitoring tool that assessed a patient's condition after a heart attack, and 4) an inpatient monitoring tool that assessed

post-surgical sepsis risk.<sup>33-36</sup> Participants were encouraged to log in on two consecutive days to answer the posted questions and comment on other participants' posted responses. All participants were given an anonymous account name by FocusGroupIt when they logged in (i.e., Participant 1, Participant 2). On both days, four members of the research team (AMS, BAB, SAM, and XZ) reviewed participants' posts throughout the day and posted follow-up questions and responses.

**Box 1.** Vignette example depicting healthcare AI used in online focus groups.

A patient comes to the emergency department with chest pain and shortness of breath. Dr. Rodriguez, an emergency medicine intern in training, takes a full history and performs a physical exam of the patient's heart and lungs. During her examination, Dr. Rodriguez uses an artificial intelligence enabled stethoscope that provides enhanced listening of the patient's heart and creates a recording for the physician to review after the examination. The artificial intelligence can detect nuances in heart sounds that may go unnoticed by inexperienced ears. After completing her examination, Dr. Rodriguez identifies a preliminary diagnosis of a heart murmur, sees the stethoscope's readings also indicate that a heart murmur is likely, and schedules cardiology appointment for the patient to follow-up.

After participants responded to all questions, they were redirected to a separate survey hosted in REDCap where they could provide their contact information to receive \$25 in remuneration.<sup>37, 38</sup> Personal information entered in REDCap could not be linked to participants' focus group responses. Participants who had agreed to participate in the focus group but had not entered information into REDCap at the end of day two, received an email reminder that the final focus group section provided a link for entering their information for remuneration. We left focus groups open for three days in case some participants needed extra time to complete their responses. Moderators did not post any follow-up prompts after the second day.

## Analysis

We applied a rapid analytic approach to the qualitative focus group data.<sup>39</sup> All free-text responses from each focus group were exported verbatim and used to create individual summary tables that synthesized content according to thematic headings that corresponded to questions posed in the group. Three study team members conducted content analyses (AMS, SAM, and XZ).<sup>40</sup> One team member generated initial summary tables, while another reviewed the tables for accuracy and supplemented content. Once completed, the summary tables were collapsed across focus groups and memoranda were generated that identified major themes. The study team discussed major themes and key findings and arrived at consensus. A thematic map illustrating patient information needs was developed based on an inductive analysis of participant responses to questions focused on AI transparency (i.e., *what would you want to know if AI was being used in your healthcare?*). Trust factors were also identified based on patient responses to questions focused on trust (i.e., *what would you need to trust the AI results or recommendations?*) as well as other responses directly referencing trust. Multiple choice question responses were aggregated, and descriptive statistics were generated using R (version 4.2.1).

## Results

Eighty-five participants were contacted via email with an invitation to participate in the study. Participants who responded (n=42) indicating an interest in participating were enrolled in a focus group.

We conducted three focus groups with a total of 42 participants between October and November of 2022. Demographic characteristics of study participants are presented in Table 1. We report on main and sub-themes related to 1) information about AI development, oversight, and impact on care as well as trust factors, 2) preference for model documentation and information delivery via labels and other mechanisms, 3) importance of disclosure of AI use when critical to provider decision making or care provision, 4) patient perceptions of consent to use AI, 5) patient preferences for deciding whether AI is used in their care, and 6) favorable perceptions of physicians using AI. In addition to these findings, we present descriptive statistics for responses to attitudinal questions concerning AI transparency detailed in Table 2.

### **Information about the AI tool, oversight, and impact on care experience influence trust**

Participants expressed information needs that generally fell into one of several major and sub-themes including information about the tool, oversight mechanisms, and impact on care as illustrated in a thematic map in Figure 1.

#### ***Information about the AI tool***

Participants wanted to know the general characteristics of AI tools used in their care such as its name, function, and purpose. Participants also wanted to know who manufactured the tool, which influenced their trust of a given system due to perceptions of reliability.

*I would want to know the company that produced the software in the smart watch and the company that produced the smart watch. Their reliability tells more about the accuracy of the software. (FG3)*

Patients also sought more evidence-oriented information such as the training data used, past performance, and its generalizability. Discussions of AI performance by patients concentrated on tool accuracy, its proven effectiveness, and reliability. These factors impacted perceptions of trust and effectiveness and were relevant to concerns about algorithmic bias and errors.

*I would want to know what studies were done using the particular AI being used in my healthcare - and how reliable, effective, and what outcomes/impact on care it had. (FG3)*

*In order to trust the results, I guess I'd want to know that people like me have been represented in the tests, clinical trials, etc. (FG2)*

Participants also wanted to be informed about a tool's limitations and potential risks.

*I would want to know about any limitations needed while using it, which the provider would usually review with me anyway. Trust results if verified as correct by provider. (FG1)*

#### ***Information about oversight***

Patients wanted to know what oversight procedures were in place for any AI tools used in their care which was also a factor influencing trust. Effective oversight was described as a distributed and layered process undertaken by regulatory bodies, care teams, and even other patients.

*I would imagine a board with medical providers and patient representatives setting rules and procedures for all kinds of situations and monitoring compliance. (FG1)*

*I would feel comfortable, knowing fully well it's been approved by FDA and other relevant agencies governing it's [sic] usage. (FG3)*

Additionally, a recurring theme was patients' preference for their care teams to review and confirm the insights from AI systems used in their care, leading to greater trust.

*Trust in the AI will come when my healthcare team reviews the data collected and agrees with the AI generated recommendations. (FG1)*

Lastly, participants felt strongly that there should be transparent and reliable oversight mechanisms in place for the use and storage of their data. They wanted to know what data stewardship practices were in place with particular concern for whether their data might be shared or sold, which influenced their trust in these systems.

*I see this as a regulatory concern. I would think that AI presents a big logistical challenge with regard to sharing patient data culled from medical records. Privacy and confidentiality need to somehow be maintained. (FG1)*

### **Impact on care experience**

Participants wanted to know precisely how AI tools would be used in their care such as how AI would factor into clinical decision making, the added value AI provides, and how care involving AI compared to standard of care procedures.

*I would appreciate transparent and clear communication about the role and impact of utilizing AI in that healthcare service setting. (FG3)*

*I would want to know how it is done and comparable to an in-office test, something that was told to me before I actually got the warning to trust that it is accurate. (FG2)*

Participants also desired information on any financial costs associated with the use of AI in their care.

*I would want to know how much this [AI] is costing me. Especially if this is something you would usually have as a follow up with a medical professional. (FG2)*

### **Preference for model documentation and information delivery via labels and other mechanisms**

Most participants were receptive to the idea of labeling AI tools. This was reinforced by the multiple-choice question data, with 78.6% (n=33) participants responding that labels would be necessary for the use of AI tools in their care. Patients also emphasized that labels should be designed with an emphasis on interpretability for patients.

*Labels for AI software are a great idea, in principle. They need to be understandable to the layman. (FG1)*

However, some participants challenged the idea of an informational label suggesting that they might contribute to misunderstanding. Labels could potentially create confusion or fail to adequately communicate complex information about AI models.

*I think the struggle with this is the communication of what these labels mean. It is not common knowledge of everything that is described on nutrition labels, and I think it would be similar and could lead to confusion in the health care setting further. (FG2)*

*Labels are good, but might oversimplify the precision and huge amount of data, processing and nuances that go into any AI technology. (FG3)*

Patients also suggested other mechanisms for providing information on AI tools, including leveraging consent forms, informational videos, visit summaries, or discussions with their care team.

*I think that any information that could be included on a label is best contained in a consent form signed with your doctor for treatment of a specific condition. Since different AI software would be used for different scenarios, the label would vary accordingly. It needs to be linked to the specific condition it is being used to treat. (FG1)*

In addition to specifying *how* they wanted information delivered, participants also commented on *who* they would like to receive information from. Participants often stated wanting to be informed by their treating physician or other member of the care team.

*Practically I would love the information about this AI from an experienced practitioner. I would prefer the information be explained to me by an expert either directly or through emails. (FG1)*

Patients also discussed the appropriate timing for receiving information, with many participants wanting to be informed about the AI tool prior to or during its use for tools that are higher risk or are addressing more immediate concerns. Participants also indicated information about the AI tool used could be included in post-visit documentation such as an after visit summary.

*I would want to be notified on the spot directly from my Doctor, I would like the info presented to me both on paper in an after visit summary. I would also like it to be verbalized to me by that same Doctor. (FG2)*

### **Importance of disclosure of AI use when critical to provider decision making or care provision**

On the multiple-choice survey questions, 66.7% (n=28) participants indicated that it was very important or important to be informed by their care teams about every use of AI in their care, and 71.4% (n=30) participants responded that it was very important or important to receive information about how AI is used in their care. Focus group responses also indicated, most patients wanted to be informed about AI use to understand the impact on their care and how it might factor into a care team's decision.

*AI is a tool that can be used by the team and I would like to know how it is used. Since AI is a relatively new technology it is important that the patient is informed in order to build trust in AI. (FG1)*

Some patients noted that the context of AI use influenced their perceived importance of its disclosure; patients wanted AI to be disclosed if it played a critical role in their care (e.g., surgical procedures, diagnoses, monitoring), but were ok with not being informed about less consequential uses (e.g., scheduling, notifications).

*In small things like appointment scheduling, I think it is obvious it is AI and I do not need to be informed for it. In bigger things like device monitoring I would probably like to know. I think knowing for me is more so just because if something goes wrong, I will know who to contact or what to do. (FG3)*

Participants were mixed on being alerted to AI findings and, if so, how. Some welcomed being provided with all data gathered by AI, while others only wanted to be alerted by their doctor if AI detected something serious, warranting treatment. Some worried that notifications about minor findings could cause undue alarm.

*I think that a lot of caution is needed in presenting AI analysis results. They could cause a lot of unnecessary stress to individuals involved. I also think this need[s] ethical considerations. What if*

*an AI analysis predicts a non-curable, terminal illness? What are the ethical considerations for that kind of analysis and information. (FG1)*

### **Variability in patient perceptions of consent to use AI**

Participants expressed mixed opinions regarding the need to provide consent for AI to analyze their data. Some thought their data was likely already being used and that consent was already covered under agreement to be treated.

*I could not care less. I almost have some expectation that it [AI] will be utilized without my knowledge or consent. I don't see the need to make this any different than a doctor using any computer system to assist them. (FG2)*

Others valued consent but did not see the need for it to be provided each time AI was used. Some patients thought a broad consent would suffice; others wanted to provide consent for particular uses. For instance, participants were more concerned about consenting when their personal information might be shared with third-parties and less concerned when used by their care team for treatment.

*I would feel uneasy about this unless it was disclosed and I consented to the continuous monitoring. Once consented, similar to the waiver for all healthcare data collected at [Clinic] to be open for research purposes, I found this honest and innovative that my data may help others in the future. (FG2)*

Finally, some participants wanted to provide consent every time AI was analyzing their data.

*I would have to consent to any of this. It sounds like overkill. I would be concerned about the security of the information. (FG1)*

### **Variability in patients' preferences for deciding whether AI is used in their care**

66.7% (n=28) participants indicated that it was very important or important to be involved decision-making regarding the use of AI in their care. Furthermore, 73.8% (n=31) responded that it was important to have a provider explain how a decision was made based on an AI result. Patients expressed varied opinions on needing to be involved in decision-making concerning healthcare AI. Many felt they should be involved in the decision to use AI just as they would want to actively make decisions about other aspects of their treatment.

*My decision must be required before I can allow a machine work on me even if it's 100% accurate in its dealings and more accurate than humans. (FG3)*

Other participants envisioned leaning on the expertise of their care team, but still maintaining authority in deciding whether AI would be used.

*If the provider is supposed to be an expert, I expect them to give me their opinion about what's best. Of course, as a patient, I would always have to right to yay or nay its use. That said, I always anticipate being a partner in my healthcare decisions with any providers. (FG2)*

Some patients were comfortable deferring to their care team's judgment, suggesting that they would have more expertise to make an informed decision. However, informing patients about the use of AI was still desired.

*I'm not sure I'm qualified enough to make an informed decision about its use. I would like to know if and how it's being used but I would leave the decision up to the care team. (FG1)*

## Mostly favorable perceptions of physicians using AI

Participants generally expressed that their opinion of their provider would not change if their provider used AI in their practice. Furthermore, others trusted their physician's judgment on when AI might be a helpful tool based on the physician's clinical knowledge.

*It [AI] would not impact my trust for the experienced physician at all because ultimately, I trust that physicians use tools and experience to come up with the diagnosis. The AI is there as a technologically enhanced tool. (FG1)*

Many viewed the use of AI as a positive sign that their physician was keeping up to date with new technology and using it in conjunction with their clinical expertise to provide care and make recommendations.

*It wouldn't change my opinion in a bad way but might in a good way. For example, I love to read and am often surprised that I might have read something that a clinician has not. It makes me think that they don't try hard enough to keep up with new research. If I discovered that a provider was using AI software, I think I'd think more highly of them, depending on the purpose of the software. (FG2)*

The extent to which participants had cultivated relationships with their physicians over significant periods of time contributed to this level of trust, irrespective of the tools used.

*It would not change my opinion of my healthcare provider. How I feel about him/her is based on personal interaction, knowledge of my medical history, and the trust that has been built up over the years. (FG1)*

In circumstances where a physician disagreed with an AI recommendation, most participants would not view the physician negatively. Some patients would even view their physician more positively for not deferring to the AI recommendation.

*I would believe that the provider is thinking critically about the results of the AI's interpretation of the data it was given, and as long as the explanation and rationale is given plausibly and logically, I would be reassured by the disagreement. If it kept happening though repeatedly in a pattern, it would concern me that something systemically, on either (or both) [the] part of the prescriber or AI, is wrong. (FG3)*

However, some participants would want reassurance if their physician disagreed with the AI's recommendation and might seek out a second opinion from another physician.

*I usually do some research before even seeing a provider so I have a notion of what they'll say and even prescribe. If a doctor disagrees with an AI's result but still aligns with what I've found on my own, that would be helpful for me but I think I'd still have to seek a second opinion and let them know why I was doing that. (FG2)*

## Discussion

The present study is among the first to systematically identify information needs and factors that facilitate trust in AI from the perspective of patients. Our findings add to the growing body of work in this area by expanding, affirming, and characterizing patient information needs and trust factors identified from other specialty areas.<sup>23</sup> Furthermore, our work expands on past research demonstrating transparency is important for fostering trust in the use of AI systems in healthcare and physicians who chose to use AI.<sup>41</sup> While our findings are grounded in clinical vignettes describing AI in cardiovascular

care, it is worth noting that patients often generalized their responses to broad uses of AI in their care in addition to responding to specific vignette-based questions.

While there is consensus around the need for transparency, the best way to deliver key information that promotes transparency remains less clear. Many, both internal and external to the healthcare sector, have proposed model documentation in the form of AI labeling and sharing of associated data sets.<sup>19, 20, 42, 43</sup> Dynamic AI labeling, in particular, has been argued to be a critical part of supporting overall regulatory approaches for healthcare AI.<sup>44</sup> Our results show that while patients tend to think labeling could improve transparency, labels may not always be the best or most comprehensive approach for communicating this information. Labels instead might serve as an entry point to a broader ecosystem of information that patients can engage with according to their needs and preferences.

Patients have concerns about labels being confusing and inaccessible and therefore may prefer alternative ways of delivering this information (e.g., consent forms). This finding highlights the importance of incorporating diverse patient perspectives when designing strategies to disseminate information about AI tools and tailoring these strategies according to patient needs, preferences, and values. For instance, progressive disclosure could be used as an approach to allow patients to decide the depth of information they would like to review about AI tools used in their care, and materials could then be created accordingly (e.g., label with brief facts, patient handouts with more in depth data and information).<sup>45</sup> Visual and interactive aspects of model documentation could also be considered in support of patient needs, and further highlights the relevance of user interactions with AI systems when designing for transparency.<sup>46</sup> Additional research is needed to identify what mechanisms are most effective at delivering information and calibrating trust in AI systems and the care teams that employ them, and how to tailor delivery mechanisms to accommodate diverse patient preferences and clinical scenarios.

Our findings suggest that transparency for patients encompasses more than information availability. Patients often express a preference for being an active decision-maker in their care and use of AI in it, albeit to varying degrees. The presence of healthcare AI in clinical settings may prompt additional conversations between patients and their care teams regardless of whether labels or other delivery mechanisms were used to communicate the information. While ubiquitously stated in ethical frameworks and guidelines, transparency lacks consensus and is often complicated by the audience AI systems need to be transparent to.<sup>2, 47</sup>

Moreover, our findings suggest physicians play a pivotal role in facilitating AI transparency and mediating patient trust. Our findings suggest that patients will often look to their physician(s)' judgment on whether AI tools are safe and effective in their care, consistent with findings from other patient-centered research.<sup>48</sup> These insights emphasize a need to further understand how physicians develop trust in AI tools and how that trust may then be shared with patients.<sup>41, 49</sup> Furthermore, there are numerous implications for physicians in the adoption of AI tools, such as training, disclosure, consent, and liability that are relevant to how associated risks may be communicated to patients.<sup>12, 50-52</sup> These additional responsibilities for physicians and healthcare institutions should be evaluated as regulations are considered for AI systems.

Developing AI transparency solutions for patients fits into a broader range of strategies for enhancing overall transparency of these systems in healthcare, including for clinical and technical stakeholders. This manuscript focuses on the information and transparency needs of patients. This focus is particularly important given the dearth of research with patient populations, even though their personal data will likely be used for the development of healthcare AI systems, and they will be on the receiving end of clinical decisions driven by these tools.<sup>53</sup> By including patient perspectives in ongoing transparency work,

there are opportunities to support health equity and foster trust in healthcare AI tools.<sup>54</sup> In coordination with technical approaches, ethical frameworks, and regulatory guidelines, there is the ability to strengthen healthcare AI transparency for all stakeholders in accordance with their precise information needs.<sup>17</sup>

Our study has several limitations. Conducting focus groups in an online asynchronous format constrains the moderators' ability to ask follow-up and clarifying questions.<sup>30</sup> Some follow-up questions did not receive a response, and sometimes there was a time delay in correcting misunderstandings to probes. Participants were also limited to those who had sufficient broadband access and technological literacy to participate in an online hosted discussion.<sup>55</sup> Additionally, our findings might have been impacted by selection bias, limiting their generalizability. While ResearchMatch helped facilitate recruitment beyond a single healthcare institution and local geographic region, as an online health research-oriented database, the pool of potential participants may have possessed greater health and technological literacy than that of the general populace. Future research should further engage underrepresented populations and communities through participatory research methods to examine whether and how AI information and transparency needs differ across patient demographics, with the goal to inform AI information dissemination strategies and policies that are better tailored to varying community needs.

Participants were prompted to respond to uses of AI in healthcare broadly and more specifically through clinical vignettes based on four AI applications in cardiovascular care. However, there is immense variety in healthcare AI between numerous clinical specialties and ML methods. Thus, further research evaluating a broader range of technologies and their use in clinical specialties is necessary to better understand the nuances of information needs as they relate to specific AI applications and care circumstances. While our findings suggest that there are a set of information needs relevant for patients, our analysis is limited in its ability to generalize and characterize the relative importance of these information needs. Additional research with larger study samples is needed to better understand how these factors might be prioritized by patients for potential transparency solutions.

## **Conclusion**

As healthcare stakeholders grapple with the adoption of AI into clinical settings, consideration for patient information needs will be vital for calibration of trust and determinations about safe and effective use of these tools. The present study outlines several of those needs along with key patient perspectives on information delivery, decision-making involvement, and physician AI use. While some of these factors overlap with transparency solutions outlined for technical and clinical experts, insights from patients can aid the expansion of these approaches to better support patient-centered adoption of healthcare AI.

## **Acknowledgements**

Not applicable

## **Author contributions**

AMS and SAM contributed to data curation, formal analysis, investigation, and writing. XZ contributed to the conceptualization, methodology, data curation, formal analysis, investigation, and writing. JLR contributed to the conceptualization, methodology, and writing. JEM contributed to the conceptualization, funding acquisition, and writing. BAB contributed to the conceptualization, funding acquisition, methodology, supervision, data curation, formal analysis, investigation, and writing.

## **Ethical approval**

The Mayo Clinic Institutional Review Board approved this study (protocol # 21-012302). All participants consented verbally prior to participate in each focus group and to the distribution of their de-identified information in accordance in accordance with IRB approvals.

## Conflicting Interests

XZ offers scientific input to research studies through a contracted services agreement between Mayo Clinic and Exact Sciences. JEM receives grant funding through Yale University from the FDA and Arnold Ventures. JEM also serves on the advisory board of GalateoBio and the board of directors for Bioethics International. BAB offers scientific input to research studies through a contracted services agreement between Mayo Clinic and Anumana, Inc. BAB also serves on the Mayo Clinic Biomedical Ethics and AI Advisory Council and on the Mayo Clinic Enterprise AI Translation Advisory Board. AMS, SAM, and JLR report no competing interests.

## Funding

This research was supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award [U01FD005938] totaling \$712,431 with 100 percent funded by FDA/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the U.S. Government.

## References

1. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology* 2017; 2: 230-243. DOI: 10.1136/svn-2017-000101.
2. Jobin A, Ienca M and Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 2019; 1: 389-399. DOI: 10.1038/s42256-019-0088-2.
3. Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: A mapping review. *Soc Sci Med* 2020; 260: 113172. DOI: 10.1016/j.socscimed.2020.113172.
4. Murphy K, Di Ruggiero E, Upshur R, et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics* 2021; 22: 14. DOI: 10.1186/s12910-021-00577-8.
5. Prem E. From ethical AI frameworks to tools: a review of approaches. *AI and Ethics* 2023; 3: 699-716. DOI: 10.1007/s43681-023-00258-9.
6. Bernal J and Mazo C. Transparency of Artificial Intelligence in Healthcare: Insights from Professionals in Computing and Healthcare Worldwide. *Appl Sci-Basel* 2022; 12. DOI: 10.3390/app122010228.
7. Shick AA, Webber CM, Kiarashi N, et al. Transparency of artificial intelligence/machine learning-enabled medical devices. *npj Digital Medicine* 2024; 7: 21. DOI: 10.1038/s41746-023-00992-8.
8. Wadden JJ. Defining the undefinable: the black box problem in healthcare artificial intelligence. *J Med Ethics* 2021 20210721. DOI: 10.1136/medethics-2021-107529.
9. Whitney RL, White AEC, Rosenberg AS, et al. Trust and shared decision-making among individuals with multiple myeloma: A qualitative study. *Cancer Medicine* 2021; 10: 8040-8057. DOI: 10.1002/cam4.4322.
10. Abramoff MD, Tarver ME, Loyo-Berrios N, et al. Considerations for addressing bias in artificial intelligence for health equity. *NPJ Digit Med* 2023; 6: 170. DOI: 10.1038/s41746-023-00913-9.
11. Nundy S, Montgomery T and Wachter RM. Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence. *JAMA* 2019; 322: 497-498. DOI: 10.1001/jama.2018.20563.
12. Kiener M. Artificial intelligence in medicine and the disclosure of risks. *AI Soc* 2021; 36: 705-713. DOI: 10.1007/s00146-020-01085-w.

13. Asselbergs FW and Fraser AG. Artificial intelligence in cardiology: the debate continues. *Eur Heart J Digit Health* 2021; 2: 721-726. 20211018. DOI: 10.1093/ehjdh/ztab090.
14. Chaddad A, Peng J, Xu J, et al. Survey of Explainable AI Techniques in Healthcare. *Sensors (Basel)* 2023; 23 20230105. DOI: 10.3390/s23020634.
15. Loh HW, Ooi CP, Seoni S, et al. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). *Comput Methods Programs Biomed* 2022; 226: 107161. 20220927. DOI: 10.1016/j.cmpb.2022.107161.
16. Petch J, Di S and Nelson W. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Can J Cardiol* 2022; 38: 204-213. 20210914. DOI: 10.1016/j.cjca.2021.09.004.
17. Gerlings J, Jensen MS and Shollo A. Explainable AI, But Explainable to Whom? An Exploratory Case Study of xAI in Healthcare. In: Lim C-P, Chen Y-W, Vaidya A, et al. (eds) *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*. Cham: Springer International Publishing, 2022, pp.169-198.
18. Brereton TA, Malik MM, Lifson M, et al. The Role of Artificial Intelligence Model Documentation in Translational Science: Scoping Review. *Interact J Med Res* 2023; 12: e45903. 20230714. DOI: 10.2196/45903.
19. Mitchell M, Wu S, Zaldivar A, et al. Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*. Atlanta, GA, USA: Association for Computing Machinery, 2019, p. 220–229.
20. Sendak MP, Gao M, Brajer N, et al. Presenting machine learning model information to clinical end users with model facts labels. *NPJ Digit Med* 2020; 3: 41. 20200323. DOI: 10.1038/s41746-020-0253-3.
21. Gebru T, Morgenstern J, Vecchione B, et al. Datasheets for datasets. *Commun ACM* 2021; 64: 86–92. DOI: 10.1145/3458723.
22. Holland S, Hosny A, Newman S, et al. The dataset nutrition label. *Data Protection and Privacy* 2020; 12: 1.
23. Robinson R, Liday C, Lee S, et al. Artificial Intelligence in Health Care—Understanding Patient Information Needs and Designing Comprehensible Transparency: Qualitative Study. *JMIR AI* 2023; 2: e46487. Original Paper 19.6.2023. DOI: 10.2196/46487.
24. Nelson CA, Pérez-Chada LM, Creadore A, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. *JAMA Dermatology* 2020; 156: 501-512. DOI: 10.1001/jamadermatol.2019.5014.
25. Jutzi TB, Kriehoff-Henning EI, Holland-Letz T, et al. Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective. *Frontiers in Medicine* 2020; 7. Original Research. DOI: 10.3389/fmed.2020.00233.
26. Young AT, Amara D, Bhattacharya A, et al. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *The Lancet Digital Health* 2021; 3: e599-e611. DOI: 10.1016/S2589-7500(21)00132-1.
27. Reddy S, Allan S, Coghlan S, et al. A governance model for the application of AI in health care. *J Am Med Inform Assoc* 2020; 27: 491-497. DOI: 10.1093/jamia/ocx192.
28. Balla S, Gomez SE and Rodriguez F. Disparities in Cardiovascular Care and Outcomes for Women From Racial/Ethnic Minority Backgrounds. *Curr Treat Options Cardiovasc Med* 2020; 22: 75. 20201117. DOI: 10.1007/s11936-020-00869-z.
29. Ifidon AM, Agbalajobi O and Essien UR. Improving Racial and Ethnic Minority Representation in Cardiovascular Disease Trials to Advance Health Equity. *JAMA Cardiol* 2020; 5: 611. DOI: 10.1001/jamacardio.2020.0167.

30. Gordon AR, Calzo JP, Eiduson R, et al. Asynchronous Online Focus Groups for Health Research: Case Study and Lessons Learned. *International Journal of Qualitative Methods* 2021; 20: 1609406921990489. DOI: 10.1177/1609406921990489.
31. Hatten JE. *Optimal size for online asynchronous text-based focus group discussions: A mixed methods study*. University of Minnesota, 2014.
32. Murphy J, Hughes J, Read S, et al. Evidence and practice: a review of vignettes in qualitative research. *Nurse Res* 2021; 29: 8-14. 20210527. DOI: 10.7748/nr.2021.e1787.
33. Bumgarner JM, Lambert CT, Hussein AA, et al. Smartwatch Algorithm for Automated Detection of Atrial Fibrillation. *J Am Coll Cardiol* 2018; 71: 2381-2388. 20180310. DOI: 10.1016/j.jacc.2018.03.003.
34. Kim Y, Hyon Y, Lee S, et al. The coming era of a new auscultation system for analyzing respiratory sounds. *BMC Pulmonary Medicine* 2022; 22: 119. DOI: 10.1186/s12890-022-01896-1.
35. McElroy I, Sareh S, Zhu A, et al. Use of digital health kits to reduce readmission after cardiac surgery. *J Surg Res* 2016; 204: 1-7. 20160423. DOI: 10.1016/j.jss.2016.04.028.
36. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med* 2021; 181: 1065-1070. DOI: 10.1001/jamainternmed.2021.2626.
37. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009; 42: 377-381. DOI: 10.1016/j.jbi.2008.08.010.
38. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 2019; 95: 103208. DOI: 10.1016/j.jbi.2019.103208.
39. Nevedal AL, Reardon CM, Opra Widerquist MA, et al. Rapid versus traditional qualitative analysis using the Consolidated Framework for Implementation Research (CFIR). *Implement Sci* 2021; 16: 67. DOI: 10.1186/s13012-021-01111-5.
40. Forman J and Damschroder L. Qualitative Content Analysis. In: Jacoby L and Siminoff LA (eds) *Empirical Methods for Bioethics: A Primer*. Emerald Group Publishing Limited, 2007, pp.39-62.
41. Asan O, Bayrak AE and Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *J Med Internet Res* 2020; 22: e15154. 20200619. DOI: 10.2196/15154.
42. Seifert C, Scherzinger S and Wiese L. Towards Generating Consumer Labels for Machine Learning Models. In: *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)* 12-14 Dec. 2019 2019, pp.173-179.
43. Sun C, Asudeh A, Jagadish HV, et al. MithraLabel: Flexible Dataset Nutritional Labels for Responsible Data Science. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Beijing, China: Association for Computing Machinery, 2019, p. 2893–2896.
44. Gerke S. "Nutrition Facts Labels" for Artificial Intelligence/Machine Learning-Based Medical Devices-The Urgent Need for Labeling Standards. *Geo Wash L Rev* 2023; 91: 79.
45. Springer A and Whittaker S. Progressive Disclosure: When, Why, and How Do Users Want Algorithmic Transparency Information? *ACM Trans Interact Intell Syst* 2020; 10: Article 29. DOI: 10.1145/3374218.
46. Crisan A, Drouhard M, Vig J, et al. Interactive Model Cards: A Human-Centered Approach to Model Documentation. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. <conf-loc>, <city>Seoul</city>, <country>Republic of Korea</country>, </conf-loc>: Association for Computing Machinery, 2022, p. 427–439.
47. Stoyanovich J, Van Bavel JJ and West TV. The imperative of interpretable machines. *Nat Mach Intell* 2020; 2: 197-199. DOI: 10.1038/s42256-020-0171-8.
48. Richardson JP, Smith C, Curtis S, et al. Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ Digit Med* 2021; 4: 140. 20210921. DOI: 10.1038/s41746-021-00509-1.

49. Winter P and Carusi A. 'If You're Going to Trust the Machine, Then That Trust Has Got to Be Based on Something': : Validation and the Co-Constitution of Trust in Developing Artificial Intelligence (AI) for the Early Diagnosis of Pulmonary Hypertension (PH). *Science & Technology Studies* 2022; 35: 58-77. DOI: 10.23987/sts.102198.
50. Kassam I, Ilkina D, Kemp J, et al. Patient Perspectives and Preferences for Consent in the Digital Health Context: State-of-the-art Literature Review. *J Med Internet Res* 2023; 25: e42507. Review 10.2.2023. DOI: 10.2196/42507.
51. Rowland SP, Fitzgerald JE, Lungren M, et al. Digital health technology-specific risks for medical malpractice liability. *npj Digital Medicine* 2022; 5: 157. DOI: 10.1038/s41746-022-00698-3.
52. Park HJ. Patient perspectives on informed consent for medical AI: A web-based experiment. *DIGITAL HEALTH* 2024; 10: 20552076241247938. DOI: 10.1177/20552076241247938.
53. Celi LA, Cellini J, Charpignon M-L, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health* 2022; 1: e0000022. DOI: 10.1371/journal.pdig.0000022.
54. Banerjee S, Alsop P, Jones L, et al. Patient and public involvement to build trust in artificial intelligence: A framework, tools, and case studies. *Patterns (N Y)* 2022; 3: 100506. 20220610. DOI: 10.1016/j.patter.2022.100506.
55. Moore T, McKee K and McCoughlin P. Online focus groups and qualitative research in the social sciences: their merits and limitations in a study of housing and youth. *People, Place and Policy Onlin* 2015; 9: 17-28. DOI: 10.3351/ppp.0009.0001.0002.

## Tables and Figures

Table 1: Demographic characteristics of 42 patients who participated in focus groups.

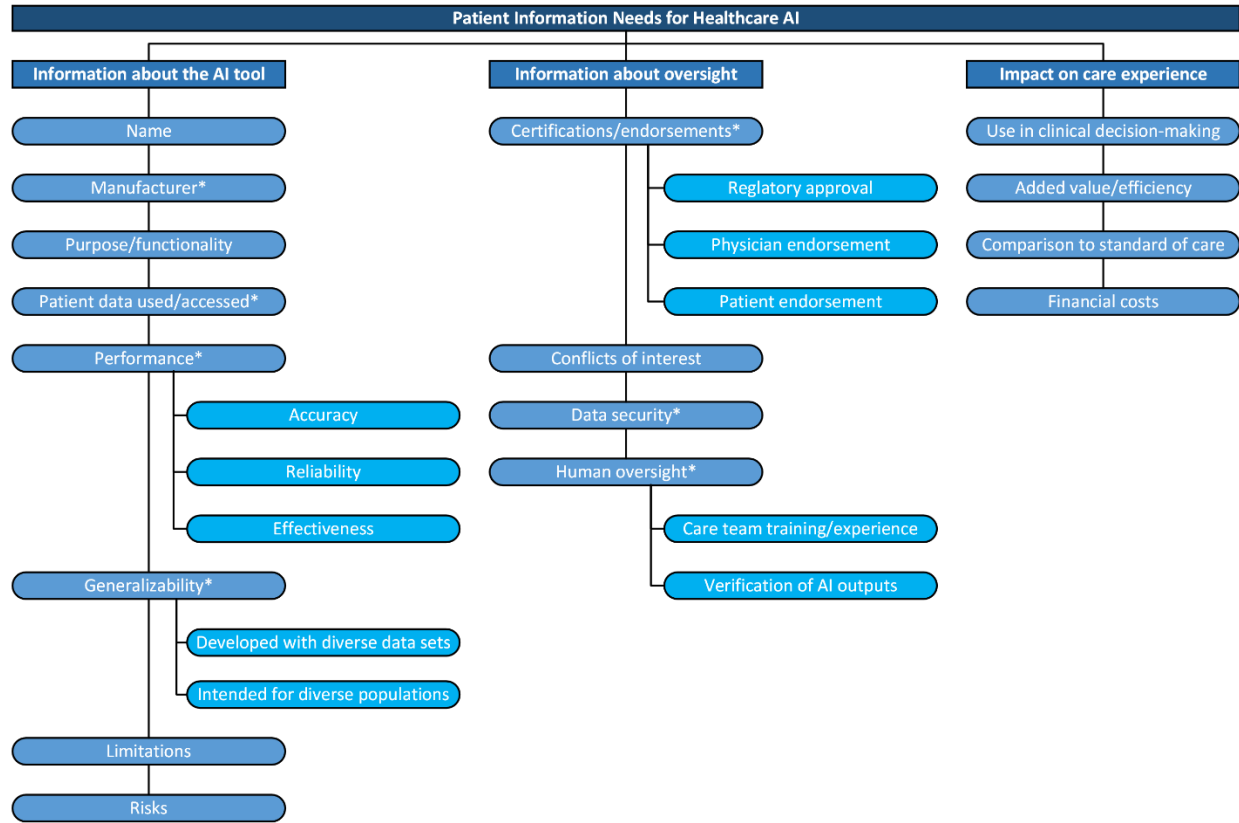
Table 2: Descriptive statistics of participant responses to attitudinal questions regarding healthcare AI transparency.

Figure 1: Thematic map of patient information needs for healthcare AI.

| <b>Table 1. Demographic characteristics of 42 patients who participated in focus groups.</b> |              |
|----------------------------------------------------------------------------------------------|--------------|
|                                                                                              | <b>N (%)</b> |
| <b>Age</b>                                                                                   |              |
| 18-24                                                                                        | 3 (7.1)      |
| 25-34                                                                                        | 11 (26.2)    |
| 35-44                                                                                        | 10 (23.8)    |
| 45-54                                                                                        | 7 (16.7)     |
| 55-64                                                                                        | 6 (14.3)     |
| 65 and above                                                                                 | 5 (11.9)     |
| <b>Gender</b>                                                                                |              |
| Man                                                                                          | 20 (47.6)    |
| Woman                                                                                        | 22 (52.4)    |
| <b>Race/Ethnicity</b>                                                                        |              |
| American Indian or Alaskan Native                                                            | 3 (7.1)      |
| Black or African American                                                                    | 15 (35.7)    |
| Hispanic/Latinx                                                                              | 1 (2.4)      |
| Asian                                                                                        | 6 (14.3)     |
| Multi-race                                                                                   | 2 (4.8)      |
| Non-Hispanic White                                                                           | 15 (35.7)    |
| <b>Education Level</b>                                                                       |              |
| High school graduate or GED completed                                                        | 2 (4.8)      |
| Completed a vocational, trade, or business school program                                    | 2 (4.8)      |
| Associate's degree (e.g., AA, AS)                                                            | 4 (9.5)      |
| Bachelor's degree (e.g., BA, BS, AB)                                                         | 15 (35.7)    |
| Master's degree (e.g., MA, MS, MSW, MBA)                                                     | 15 (35.7)    |
| Doctorate or professional degree (e.g., PhD, MD)                                             | 4 (9.5)      |
| <b>Marital Status</b>                                                                        |              |
| Married                                                                                      | 15 (35.7)    |
| Never married                                                                                | 14 (33.3)    |
| Separated, Divorced, or Widowed                                                              | 13 (31)      |
| <b>Perceived Financial Status</b>                                                            |              |
| Living comfortably on present income                                                         | 17 (40.5)    |
| Getting by on present income                                                                 | 21 (50)      |
| Finding it difficult on present income                                                       | 0 (0)        |
| Finding it very difficult on present income                                                  | 3 (7.1)      |
| Prefer not to answer                                                                         | 1 (2.4)      |

| <b>Table 2.</b> Descriptive statistics of participant responses to attitudinal questions regarding healthcare AI transparency. |              |
|--------------------------------------------------------------------------------------------------------------------------------|--------------|
|                                                                                                                                | <b>N (%)</b> |
| <b>Importance of being informed by care team about every AI use in your care</b>                                               |              |
| Very important/Important                                                                                                       | 28 (66.7)    |
| Neutral/Unimportant/Very unimportant                                                                                           | 14 (33.3)    |
| <b>Importance of participating in decision-making regarding use of AI in your care</b>                                         |              |
| Very important/Important                                                                                                       | 28 (66.7)    |
| Neutral/Unimportant/Very unimportant                                                                                           | 14 (33.3)    |
| <b>Importance of getting information about how AI is used in your care</b>                                                     |              |
| Very important/Important                                                                                                       | 30 (71.4)    |
| Neutral/Unimportant/Very unimportant                                                                                           | 12 (28.6)    |
| <b>Do you think labels for AI software are needed?</b>                                                                         |              |
| Yes                                                                                                                            | 33 (78.6)    |
| No                                                                                                                             | 9 (21.4)     |
| <b>Importance of provider explaining how they made a decision about your care based on AI result</b>                           |              |
| Very important/Important                                                                                                       | 31 (73.8)    |
| Neutral/Unimportant/Very unimportant                                                                                           | 11 (26.2)    |

**Figure 1.** Thematic map of patient information needs for healthcare AI.



High-level categories of information needs identified from inductive analysis include information about the AI tool, information about oversight, and the impact on the care experience. Information factors are enumerated for each category. Asterisks indicate factors that mediate trust.