JAMIA Open, 2026, **9(2)**, ooag029
https://doi.org/10.1093/jamiaopen/ooag029
Brief Communications

Brief Communications

Patient perspectives about deployment of artificial intelligence decision support tools in a safety-net healthcare system

Nicholas Phelps, MD[1], Patrick Vossler, PhD[2,3,4], Avni Kothari, MS[1,3,4], Katherine Steineman, BA[1,3,4], Seth Goldman, MD[1,3,4], Arturo Gasga, MD[1,3,4], Susan Ehrlich, MD, MPP[3,4], Hemal Kanzaria, MD[3,4,5], Julia Adler-Milstein, PhD[1], Jean Feng, PhD, MS[2,3,4], Paige Nong, PhD[6], Lucas S. Zier, MD, MS[1,3,4,*]

[1]Department of Medicine, University of California San Francisco, San Francisco, CA 94122, United States, [2]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA 94122, United States, [3]Zuckerberg San Francisco General Hospital, San Francisco, CA 94110, United States, [4]San Francisco Department of Public Health, San Francisco, CA 94102, United States, [5]Department of Emergency Medicine, University of California San Francisco, San Francisco, CA 94122, United States, [6]Division of Health Policy and Management, University of Minnesota Twin Cities, Minneapolis, MN 55455, United States

*Corresponding author: Lucas S. Zier, MD, MS, University of California San Francisco Department of Medicine, Division of Cardiology; Zuckerberg San Francisco General Hospital, 1001 Potrero Avenue, Room 5G2, San Francisco, CA 94110, United States (Lucas.zier@ucsf.edu)

Abstract

Objectives: To assess patient awareness, trust, perceived benefits, and risks of artificial intelligence (AI) in clinical care within an urban safety-net health system.

Materials and Methods: We surveyed 313 patients from November 2024 to January 2025 regarding AI awareness, trust in AI-assisted decision-making, and preferences for transparency and oversight. Quantitative analyses assessed associations between AI awareness and perceived benefit; qualitative analysis identified themes influencing trust.

Results: While 84% were familiar with commercial AI, fewer than half recognized the use of AI in medical decision support. Greater AI awareness was associated with higher perceived benefit (all $P < .001$). Participants emphasized transparency (92%), clinician oversight (82%), and validation as critical to trust.

Discussion: This study provides one of the first assessments of patient perspectives on AI within a safety-net healthcare setting. Patients view clinical AI favorably but demand transparency and clinician involvement.

Conclusions: Patient education and engagement are essential for equitable, trustworthy AI deployment.

Lay Summary

What We Wanted to Find Out: We wanted to learn what patients at a city hospital for people with lower incomes think about artificial intelligence (AI) in their healthcare. We looked at what they knew about AI, if they trusted it, and what good or bad things they thought it might bring.
How We Did It: We asked 313 patients between November 2024 and January 2025 to fill out a survey. We asked them about their knowledge of AI, if they would trust AI to help doctors make decisions, and if they wanted things to be clear and supervised. We used numbers to see if knowing more about AI made them think it was more helpful. We also read their written comments to understand what built their trust.
What We Found: Most patients (84%) knew about AI from everyday things like phones or smart speakers. But less than half knew about AI being used in healthcare to help doctors. Patients who knew more about AI also thought it was more helpful. Most patients said they needed AI to be clear (92%), doctors to check it (82%), and for it to be proven safe and effective to trust it.
What This Means: This study is among the first to examine how patients at a safety-net hospital feel about AI in healthcare. Patients generally like the idea of AI helping doctors, but they really want it to be clear how AI is used and for doctors to still be in charge.
Our Main Point: To make sure AI in healthcare is fair and trustworthy for everyone, it's very important to teach patients about it and involve them in how it's used.

Key words: artificial intelligence; predictive methods; generative artificial intelligence; patient attitudes.

Introduction

Artificial intelligence (AI) holds immense promise to improve health outcomes and advance equity, particularly in resource-limited healthcare settings.[1] Safety-net healthcare systems, which serve vulnerable populations, are increasingly adopting AI-driven decision-support tools to enhance clinical care. However, little is known about how patients perceive the deployment of these technologies, including their expectations, concerns, and trust in AI-assisted decision-making.[1,2] While previous work has examined clinician perspectives,[3,4,5]

Received: October 31, 2025; **Revised:** January 27, 2026; **Accepted:** February 23, 2026

patient voices remain underrepresented in discussions about AI governance, particularly within marginalized communities. We sought to assess patient perceptions of AI within our urban safety-net healthcare system. Specifically, we explored patient awareness of AI applications in healthcare, their trust in AI-assisted recommendations, and their expectations regarding transparency, clinician oversight, and performance validation. As AI becomes more integrated into clinical workflows, patient education may play a pivotal role in shaping trust and adoption. Here, we present the results of a survey designed to inform strategies for responsible AI deployment in the safety-net setting.

Methods

From November 2024 to January 2025, we distributed an electronic survey to patients at an urban safety-net healthcare system. The survey included a validated technology literacy assessment,[6] questions on AI in healthcare, and 3 AI-enabled decision-support scenarios: predictive AI for acute illness, predictive AI for chronic illness, and generative AI for chronic illness management. Descriptive statistics summarized responses. A chi-square test compared perceived accuracy and trust between predictive and generative AI. Exploratory factor analysis revealed distinct patterns in AI awareness, with factor loadings suggesting that participants' familiarity with AI applications clustered into separate dimensions (general commercial AI, predictive clinical AI, and generative clinical AI). Participants were therefore stratified by their overall AI awareness level to examine how familiarity with AI technologies mediated perceptions of accuracy and trust. Effect sizes were calculated using epsilon-squared (ε^2) for the Kruskal–Wallis H-tests to quantify the magnitude of differences across AI awareness levels. The authors performed a thematic analysis of free-text responses using a codebook approach, with 3 members of the research team independently coding them. Discrepancies were reconciled in a synchronous meeting, and final themes were agreed upon by all 3 coders. These themes were operationalized in a structured prompt for LLM-assisted annotation using GPT-4o-mini via a HIPAA-compliant API (temperature = 0.0, fixed random seed). The LLM assigned the single most appropriate theme to each of 939 responses and provided a justification for each annotation; responses not fitting any theme were coded as "None." Investigators reviewed all "None" annotations, which primarily consisted of brief non-substantive responses (eg, "I don't know"). Theme counts were then tallied from the LLM annotations. Analysis was conducted in July 2025; additional technical details are provided in the Supplement.

Results

A total of 313 individuals (response rate 5%) completed the survey. Responses were inverse weighted by survey completion rate to reflect the patient characteristics of the entire hospital. See Table 1 for patient characteristics of survey responders before and after reweighting of the data. Technology literacy was high, with 71% having high tech literacy. In contrast, there was a substantial difference in awareness between medical and non-medical AI uses: 84% were familiar with commercial applications, but fewer than half knew of clinical use in medicine—43% had heard of AI for disease diagnosis, 40% for predicting illness risk or mortality, and 47% for summarizing medical records. Participants were stratified into low and high AI awareness based on self-reported familiarity with AI in general and knowledge of specific AI applications.

AI transparency was a dominant concern, with 92% wanting notification when AI tools are used in their clinical care. Additionally, 82% emphasized "human-in-the-loop" systems,

Table 1. Characteristics of participants

Characteristic	Unweighted	Weighted	ZSFG population (%)
Age—no. (%)			
16-24	5 (2)	10 (2)	1
25-34	28 (9)	67 (11)	8
35-44	45 (15)	97 (16)	16
45-54	59 (19)	123 (20)	19
55-64	87 (28)	163 (26)	25
65-74	66 (21)	117 (19)	21
75-84	16 (5)	28 (5)	7
85-94	2 (1)	6 (1)	2
95+			0
Prefer not to say	2 (1)	4 (1)	0
Gender—no. (%)			
Male	152 (49)	297 (48)	45
Female	146 (47)	298 (48)	55
Transgender	4 (1)	7 (1)	
Nonbinary/gender-nonconforming	6 (2)	9 (1)	
Prefer not to say	2 (1)	4 (1)	0
Race/ethnicity—no. (%)			
American Indian or Alaska Native	15 (5)	30 (5)	1
Asian	35 (11)	92 (15)	17
Black or African American	44 (14)	77 (13)	11
Native Hawaiian or Other Pacific Islander	5 (2)	14 (2)	1
White	151 (49)	244 (40)	22
Other	94 (30)	224 (36)	5
Hispanic	72 (23)	182 (30)	42
Decline to answer			1
Race not specified			0

advocating for clinician oversight. When asked about different clinical AI use cases, participants broadly felt that the potential benefits of AI use outweighed its risks. This trend was most pronounced amongst participants with high AI awareness. When presented with specific clinical scenarios, approximately 64% of participants indicated that they would trust AI-generated recommendations given to their doctors, whether for predictive purposes (such as preventing hospital readmissions or detecting early sepsis) or for treatment guidance (such as adjusting diabetes medications or addressing substance use). When asked about risks versus benefits, participants with high AI awareness consistently perceived greater benefits relative to risks across all clinical scenarios (Kruskal–Wallis H-test, all $P < .001$). Effect sizes demonstrated medium to large

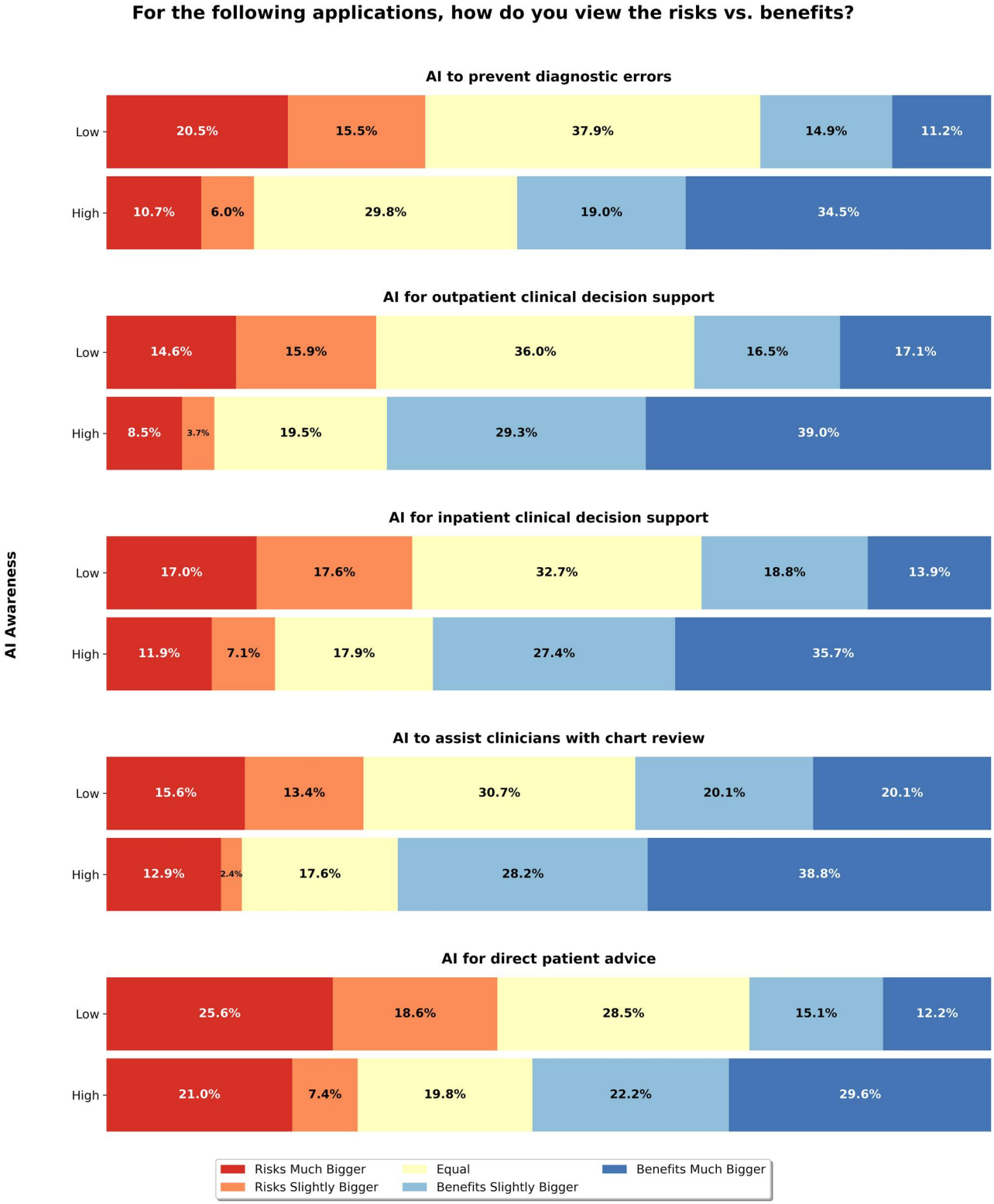

Figure 1. Patient perceptions of the risks vs benefits of the 5 clinical AI use cases.

associations between AI awareness and benefit perception, with the strongest effects observed for AI in outpatient care ($\varepsilon^2 = 0.178$) and overall healthcare applications ($\varepsilon^2 = 0.155$). Medium effects were found for AI in hospital settings ($\varepsilon^2 = 0.133$), medical records ($\varepsilon^2 = 0.088$), and patient advice ($\varepsilon^2 = 0.095$). Figure 1 shows perceptions of risk vs benefit of each clinical AI use case, stratified by AI awareness. Stratification by gender, age, and race/ethnicity revealed significant but smaller associations with risk-benefit perceptions compared to AI awareness (Table S1). Gender showed the strongest demographic effect, with men perceiving greater benefits than women across most scenarios.

Qualitative analysis identified 5 themes shaping trust in AI deployment: (1) the necessity of physician oversight, (2) demands for transparency in model performance and safety, (3) a preference to preserve human interaction in AI-assisted care, (4) interest in observing use of AI tools in clinical practice, and (5) desire for rigorous evidence of AI tools' clinical performance (see Figure 2 for a description of themes and direct quotations). Figure 2 shows the incidence of each trust theme by AI clinical scenario queried. Physician oversight and rigorous evidence were the most cited themes, particularly for generative AI in chronic disease management and AI risk prediction for chronic disease.

Discussion

This study provides one of the first assessments of patient perspectives on AI in clinical medicine within a safety-net healthcare setting. Participants were asked about real-world clinical applications of AI currently in use within our health system. Despite high general technology literacy, patient awareness of AI's clinical applications lagged behind familiarity with commercial AI. Participants were asked about real-world clinical applications of AI (including some currently in use within our health system); results are therefore applicable to the iterative improvement of these tools as well as the deployment of similar tools in the future. Participants emphasized transparency, clinician oversight, and robust validation as critical factors for fostering trust in deployed decision support tools.

While we hypothesized that patient perspectives on AI in safety-net settings might reveal distinct concerns due to unique patient populations and resource constraints, our findings largely mirror those from general populations regarding the importance of transparency, clinician oversight, and validation for fostering trust.[2,3,7,8] This suggests that foundational trust principles are broadly maintained amongst disparate patient populations and practice settings, yet their effective implementation in safety-net communities demands reinforced education and engagement to ensure equitable adoption and address existing disparities.

Notably, participants with greater AI awareness were more likely to perceive the benefits of AI as outweighing the risks, underscoring a potential role for patient education in **fostering trust in the use of these tools.** Given that AI-driven decision support is increasingly shaping patient care, efforts to educate patients on AI's capabilities, limitations, and validation processes could improve acceptance and trust.

This study has several important limitations. The low 5% response rate, despite inverse weighting to align with hospital demographics, introduces the potential for non-response bias

Theme	AI risk prediction for acute disease	AI risk prediction for chronic disease	Generative AI for chronic disease	Representative Participant Quotes
1. Physician Oversight Importance of maintaining active clinician oversight throughout use, rather than limiting involvement to initial approval	42	43	79	*"[I would trust AI more] if it was not completely autonomous...suggestions are always reviewed and approved by clinician before any application"* *"[I would want to] have the doctor audit/review the recommendations before releasing to patients"*
2. Desire for rigorous evidence Need for extended track record of AI tools' clinical performance, accuracy, and efficacy	40	53	56	*"Only documented case percentages of accurate predictions, over a wide sample and a long period of time could convince me."* *"[I would trust AI more] if I had data/research that showed this was successful."*
3. Transparency in model performance Clarity on training data, development, validation, recommendations, and differences from human decision-making	30	32	28	*"If I can understand how the AI arrives at its suggestions; what data it uses, how it assesses risks, and what algorithms it employs..."* *"The recommendations would be more trustworthy if it specifies which factors (medical history, lab tests, etc.) were taken into consideration"*
4. Preserve human interaction Maintaining doctor-patient relationships and face-to-face time with clinicians when using AI tools	32	29	20	*"[I would have more trust] 'if the AI suggestions come from the doctor's mouth.'"* *"The doctor needs to use their EQ, their intuition based on LIVED experience, in conjunction with...AI"* *"[I would trust AI more if there were] an in person review by my doctor about recommendations."*
5. Observing AI in practice Desire to observe real-world use of clinical AI tools before routine use in care	7	6	9	*"I think the more AI is used the more I would trust it."* *"[I need to become more] 'familiar with AI. What is AI, what does it look like, how does it sound and feel like?'"* *"[I feel that we] 'need more time with AI'"*

Count: 0 40 80

Figure 2. Qualitative themes (y-axis) underlying participants' trust in specific clinical AI use cases (x-axis): boxed numbers represent the incidence of each theme in participants' free-text responses regarding their trust in each specific use case.

and may limit the generalizability of our findings. Moreover, while weighting mitigated initial imbalances, our sample still exhibited an overrepresentation of White participants and an underrepresentation of Hispanic participants, suggesting that more targeted recruitment strategies are needed in future research to achieve truly representative patient populations.

These findings highlight the need for patient voices to be actively incorporated into AI policy and governance frameworks, particularly in safety-net settings where AI tools may play a crucial role in addressing healthcare disparities. Patients in our study asked reasonable and insightful questions about AI performance and validation, suggesting a need for ongoing engagement between healthcare institutions and the communities they serve. Our survey results are now being used to inform AI deployment strategies within our health system, with a focus on transparency, education, and patient-centered oversight. As AI continues to evolve in clinical medicine, ensuring that patient perspectives shape its implementation will be essential for fostering trust and equitable adoption.

Acknowledgments

The authors would like to thank the University of California, San Francisco, and the San Francisco Department of Public Health.

Author contributions

Nicholas Phelps (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Visualization, Writing—original draft, Writing—review & editing), Patrick Vossler (Data curation, Formal analysis, Investigation, Methodology, Resources, Writing—review & editing), Avni Kothari (Conceptualization, Data curation, Formal analysis, Methodology, Writing—review & editing), Katherine Steineman (Conceptualization, Investigation, Project administration, Writing—review & editing), Seth Goldman (Conceptualization, Methodology, Supervision, Writing—review & editing), Arturo Gasga (Methodology, Project administration, Resources, Writing—review & editing), Susan Ehrlich (Writing—review & editing), Hemal Kanzaria (Writing—review & editing), Julia Adler-Milstein (Data curation, Methodology, Resources, Writing—review & editing), Jean Feng (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Writing—review & editing), Paige Nong (Conceptualization, Data curation, Formal analysis, Methodology, Writing—review & editing), and Lucas S. Zier (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Writing—review & editing)

Supplementary material

Supplementary material is available at *JAMIA Open* online.

Funding

Funding for this study was provided from discretional funds of the corresponding author not associated with any existing grants.

Conflicts of interest

The authors have no conflicts of interest to disclose.

Data availability

The data underlying this article cannot be shared publicly, to maintain the privacy of individual subjects. The data will be shared on reasonable request to the corresponding author.

Human ethics and consent to participate declarations

This study was performed in accordance with the Declaration of Helsinki. All human participants provided informed consent to participate in this study. Consent was obtained in accordance with the guidelines outlined by the relevant Institutional Review Boards: the UCSF Human Research Protection Program and Institutional Review Board (approval #24-41049), and the UCSF Research at San Francisco Department of Public Health Institutional Review Board (approval #24-41049), as well as applicable ethical standards.

References

1. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25:24-29. https://doi.org/10.1038/s41591-018-0316-z
2. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *Lancet Digit Health*. 2021;3:e599-e611. https://doi.org/10.1016/S2589-7500(21)00132-1
3. Giddings R, Joseph A, Callender T, et al. Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review. *Lancet Digit Health*. 2024;6:e131-e144. https://doi.org/10.1016/S2589-7500(23)00241-8
4. Shamszare H, Choudhury A. Clinicians' perceptions of artificial intelligence: focus on workload, risk, trust, clinical decision making, and clinical integration. *Healthcare (Basel)*. 2023;11:2308. https://doi.org/10.3390/healthcare11162308
5. Fazakarley CA, Breen M, Leeson P, Thompson B, Williamson V. Experiences of using artificial intelligence in healthcare: a qualitative study of UK clinician and key stakeholder perspectives. *BMJ Open*. 2023;13:e076950. https://doi.org/10.1136/bmjopen-2023-076950
6. Nelson L, Pennings J, Sommer E, Popescu F, Barkin S. A 3-item measure of digital health care literacy: development and validation study. *JMIR Form Res*. 2022;6:e36043. https://doi.org/10.2196/36043
7. Moy S, Irannejad M, Manning SJ, et al. Patient perspectives on the use of artificial intelligence in health care: a scoping review. *J Patient Cent Res Rev*. 2024;11:51-62. https://doi.org/10.17294/2330-0698.2029
8. Esmaeilzadeh P, Mirzaei T, Dharanikota S. Patients' perceptions toward human-artificial intelligence interaction in health care: experimental study. *J Med Internet Res*. 2021;23:e25856. https://doi.org/10.2196/25856

JAMIA Open, 2026, 9, 1–5
https://doi.org/10.1093/jamiaopen/ooag029
Brief Communications