

## **DIVERSE PATIENTS' ATTITUDES TOWARDS ARTIFICIAL INTELLIGENCE (AI) IN DIAGNOSIS**

Christopher Robertson, J.D., Ph.D.<sup>1,2</sup>, Andrew Woods, J.D., Ph.D.<sup>1</sup>, Kelly Bergstrand, Ph.D.<sup>3</sup>, Jess Findley, J.D., Ph.D.<sup>1</sup>, Cayley Balsler, M.Ed., Ed.S.<sup>1</sup>, and Marvin J. Slepian, M.D.<sup>1</sup>

1. University of Arizona
2. Boston University
3. University of Texas at Arlington

**Correspondence:** Robertson, [ctr00@BU.edu](mailto:ctr00@BU.edu).

**FUNDING:** Supported by National Heart, Lung, and Blood Institute (NIH) grant 3R25HL126140-05S1.

## ABSTRACT

**Importance:** Artificial intelligence (AI) has the potential to improve diagnostic accuracy and treatment efficacy. Yet people are often reluctant to trust automated systems, and some patient populations may be particularly distrusting of healthcare innovations.

**Objective:** To determine how diverse patient populations feel about the use of AI diagnostic tools, and whether framing the choice around proven accuracy, lack of bias, or other factors affects uptake.

**Design:** After structured interviews with diverse actual patients, a randomized, blinded survey experiment placed respondents as mock patients into clinical vignettes and manipulated eight variables in factorial design. Each respondent viewed one vignette where their primary care physician (PCP) presented the choice of referral to a specialist clinic operated by an AI or human physician.

**Setting:** Internet-based survey experiment.

**Participants:** A major survey firm (YouGov) provided n=2675 responses, oversampling Black, Hispanic, Asian, and Native American populations, with reweighting to be representative of the U.S. population.

**Interventions:** The clinical vignettes were randomly manipulated in eight variables with two levels each: disease severity (leukemia versus sleep apnea), whether AI is proven more accurate than human specialists, whether the AI clinic is personalized to the patient through listening and/or tailoring, whether the AI clinic avoids racial and/or financial biases, whether the PCP promises to explain and incorporate the advice into an ultimate recommendation, and whether the PCP nudges the patient towards AI, framing it as the established, recommended, and easy choice. We also counterbalanced whether respondents answered attitudinal questions before (or after) viewing and responding to the vignettes.

**Main Outcome Measure:** Selection of AI clinic or human physician specialist clinic (binary, "AI uptake").

**Results:** With weighting representative to the U.S. population, respondents were almost evenly split (52.9% chose human doctor and 47.1% chose AI clinic). In unweighted experimental contrasts of respondents who met pre-registered criteria for engagement with the vignette, a PCP's explanation that AI has proven superior accuracy increased uptake (OR=1.48, CI 1.24-1.77,  $p<.001$ ). A PCP's nudge towards AI as the established choice also increased uptake (OR=1.25, CI: 1.05-1.50,  $p=.013$ ), as did reassurance that the AI clinic had trained counselors to listen to the patient's unique perspectives (OR=1.27, CI: 1.07-1.52,  $p=.008$ ). Disease severity (leukemia versus sleep apnea) and other manipulations did not affect AI uptake significantly. Compared to White respondents, Black respondents selected AI less often (OR=.73, CI: .55-.96,  $p=.023$ ) and Native Americans selected it more often (OR: 1.37, CI: 1.01-1.87,  $p=.041$ ). Older respondents were less likely to choose AI (OR: .99, CI: .987-.999,  $p=.03$ ), and for each unit increase in education, the odds are 1.10 greater for selecting an AI provider (OR: 1.10, CI: 1.03-1.18,  $p=.004$ ). Respondents who identified as politically conservative (OR: .65, CI: .52-.81,  $p<.001$ ) or viewed religion as important (OR: .64, CI: .52-.77,  $p<.001$ ) had significantly lower odds of choosing AI.

**Conclusions and Relevance:** Many patients appear resistant to the use of AI in diagnosis. Nonetheless information that AI systems are proven more accurate, along with nudging by caregivers and a listening patient experience, may help increase acceptance. To ensure that the benefits of AI are secured in

clinical practice, future research on best methods of physician incorporation and patient decision making is required.

**Registration:** Open Science Framework, [osf.io/9y26x](https://osf.io/9y26x).

## KEY POINTS

**Question:** How do patients feel about AI systems being used for diagnosis, and how can that decision be optimally framed to yield choices that minimize mortality and morbidity?

**Findings:** In this randomized survey experiment enriched for racial diversity but also weighted to the U.S. population, respondents expressed resistance to AI diagnosis. Proof that it was more accurate, a physician's nudge towards it as established care, and including an opportunity to talk with trained counselors to tailor treatment recommendations all supported AI uptake.

**Meaning:** Patient resistance to AI diagnosis may impinge uptake in ways that undermine treatment goals, but human physicians can support adoption, where the technology is designed with the patient experience in mind and supported by evidence of accuracy.

## INTRODUCTION

Artificial intelligence (AI) is poised to transform healthcare. Today, AI is used to analyze tumors in chest images (1), regulate implanted devices (2), and select personalized courses of care (3). Despite the promise of AI, there is broad public skepticism about AI in a range of domains from transportation to criminal justice to healthcare (4,5). Doctors and patients tend to primarily rely on doctors' clinical judgment, even when it is at odds with statistical judgment (6).

Research shows that patients prefer human doctors to AI-powered machines in diagnosis, screening, and treatment (7-10). In an early study, patients were more likely to follow medical advice from a physician than a computer and were less trustful of computers as providers of medical advice (7). Other work shows that patients are less trusting of doctors that rely on non-human decision aids (8, 9). More recently, in a series of studies, patients were less willing to schedule an appointment to be diagnosed by a robot, and they were willing to pay significantly more money for a human provider, with a reported perception that AI providers are less able to account for patients' unique characteristics (10).

Yet acceptance of AI may depend on specific features of the system and how the choice is framed, and there may be differences among groups of patients (11). Outside of healthcare, consumers have been shown to more often trust AI systems for objective tasks, while subjective tasks are viewed as more appropriate for humans (12). Some qualitative research suggests that lower levels of patient education are associated with lower trust in computerization (13). Other small studies suggest that AI may be acceptable if human physicians ultimately remain in control (14). And, although patients may prefer human physicians, some work suggests that they may better adhere to advice coming from algorithms (15).

More generally, research suggests that patients' trust in their physicians is an essential component of effective healing (16, 17). Black, Hispanic, and Native Americans reportedly have lower levels of trust in their physicians (18, 19, 20). Yet trust can be enhanced when patients and their providers have similar background, geography, or ethnic groups (21, 22, 23). Patients may be concerned about human physicians being biased by their financial relationships with pharmaceutical companies (24) or biased by implicit racial stereotypes (25). Although initial forays into algorithmic decisions gave rise to similar concerns (26), AI systems may be rigorously designed, tested, and continuously monitored to minimize racial or financial biases in healthcare (27, 28).

There are many drivers to the development and uptake of AI in healthcare, including commercial incentives and physician attitudes (29, 30). However, to the extent that patients retain the right of informed consent with regard to their own healthcare, much will depend on patient attitudes towards AI. To that end, we used qualitative and quantitative methods to study diverse patient populations' views about AI in medicine.

## METHODS

We conducted two study phases, one qualitative and one quantitative. In the qualitative phase (February to December 2020), we conducted structured interviews with 24 patients recruited for racial and ethnic diversity to understand their reactions to current and future AI technologies. In the quantitative phase (January and February 2021), we used an internet-based survey experiment oversampling Black, Hispanic, Asian, and Native American populations. Both phases placed respondents as mock patients into clinical

vignettes to explore whether they would prefer to have an AI system versus a doctor for diagnosis and treatment and under what circumstances.

### ***Development and Oversight of the Survey***

To develop clinical vignettes, we consulted physicians specializing in cardiology, pulmonology, hematology, and sleep medicine to develop vignettes, which were reviewed by physician co-author (MJS), for authenticity. This study was determined to be exempt by the Human Subjects Protection Program (Institutional Review Board) at the University of Arizona, and all subjects consented.

### ***Qualitative Pre-Study***

Our qualitative study pre-tested the vignettes and generated hypotheses. For 30-60 minute qualitative interviews in Spanish or English, we recruited 24 actual patients from clinics in Tucson, Arizona, including 10 Whites, 8 Hispanics, 3 Blacks, 2 Native Americans, and 1 Asian. 16 were females and 8 were males. Ages ranged from 19 to 92 years, with most over 50. Educational achievement was relatively high, with 7 of the subjects having a graduate degree, and 5 having a bachelor's degree.

Overall, patients expressed confidence that AI systems could achieve greater accuracy in diagnosis and treatment compared to physician. Nonetheless, several patients called on their own experiences with technology to suggest that an AI system could be fallible. Other patients, especially those who were non-White, expressed lack of trust with the healthcare system more generally and recounted anecdotes where they felt unheard or mistreated. Patients nearly uniformly said they would rely heavily on their physicians to guide their choice of whether an AI system would be used for their diagnosis or treatment, but most nonetheless emphasized that they would generally want to know of such use, suggesting that it is material for their informed consent. Several patients indicated that they had greater confidence in their human physicians than an AI system to personalize treatment decisions to the patient's own situation. The technology was more attractive for younger and more educated patients. Several patients invoked their belief in God as being important to their healthcare decisions.

### ***Quantitative Survey Experimental Design and Materials***

We designed our quantitative phase as a blinded, randomized survey experiment in factorial design, manipulating eight variables by substituting text in or out of the base vignette. Respondents were block randomized by race/ethnicity to experimental conditions. We counterbalanced whether respondents answered certain covariate questions before or after our vignettes and primary outcomes.

The full text of the vignettes and manipulations are shown in the Supplemental Materials. The base clinical vignette was split into two parts, with an initial segment laying out the patient's history and primary care physician's (PCP's) initial impressions. As one of the experimental manipulations, all respondents saw either a leukemia or a sleep apnea version of the base case, with the PCP explaining that leukemia could be fatal if not properly diagnosed and treated, while sleep apnea was described as interfering with the patient's comfort and lifestyle. Respondents were asked to explain their "reactions and feelings at this point in the story."

A final segment explained that, "your doctor would like to get a second opinion on whether you have leukemia, and if so get the best treatment plan," or "to determine whether you actually have an apnea, determine its type, and determine the best course of treatment, your physician suggests a sleep study."

The PCP then presented the choice of AI versus physician specialist, followed by the experimental manipulations. Table 1A displays the description of the two providers. Table 1B shows a summary of the manipulations, which followed this presentation.

Our primary outcome variable (“AI uptake”) was binary, “Which provider would you choose to diagnose your health problem? Dr. Williams, the specialist physician [or] The Med-X clinic, the AI computer system,” with presentation order randomized. Finally, we presented several debriefing questions. These included an attention-check question to test whether the respondent could identify the disease featured in the vignette read a few minutes prior and a self-assessment of the respondent’s understanding of the vignette, on a ten-point scale.

**\*\*Table 1 About Here\*\***

### ***Survey Administration and Screening***

YouGov interviewed 2875 U.S. adults who identified as White, Black, Latino, Asian or Native American, who were then matched down to a sample of 2675 to produce the final dataset. To allow well-powered estimates for particular racial and ethnic groups, the sample was designed to over-represent non-White Americans as follows: 650 Whites (24%), 575 Blacks (21%), 550 Latinos (21%), 550 Asians (21%), and 350 Native Americans (13%). YouGov’s matching and weighting approach is described in the Supplemental Materials.

We report both weighted and unweighted analyses. This allows us to both represent the United States population as well as investigate differences by racial groups and those who fully engaged with the vignettes. For the unweighted analyses we exclude respondents based on some of the previously established criteria in the pre-registration plan: those that failed an attention check about whether respondent could correctly recall the disease presented in the vignette, and those reporting that they did not understand the vignette (bottom two levels on 10-point scale). After these removals, the sample for our primary analyses was N=2,472. We present models without exclusions in the Supplemental Materials.

### ***Characteristics of the Respondents***

Our sample was diverse and representative after weights are applied. As shown in Table 2, in addition to race/ethnicity coverage, respondents span multiple educational levels, with roughly equal amounts of respondents in the categories of high school degree or less, some college or a two-year degree, and a four-year college degree or higher. We have slightly more females in the sample. The average age was 48 years, and about half made less than \$50,000 a year (N=1194; 54%).

**\*\*Table 2 About Here\*\***

### ***Statistical Analysis***

Our primary analyses rely on multivariable logistic regressions in STATA, and the computation of predictive margins and 95% confidence intervals (CI), exploiting the randomized design. We use  $\alpha=.05$  as the threshold for significance of p-values.

In Table 3, Model 1 merely shows effects of our experimental conditions controlling for whether covariates were collected before or after the vignettes (order) and familiarity or experience with the

illness. Model 2 adds demographic controls shown in the Table, although the sample size is smaller primarily due to missing income data. Model 3 also controls for respondents' attitudes regarding trust in providers, hospitals and AI companies, and these estimates and p-values are displayed in Figure 2 and discussed in the text, except where otherwise noted. Wald tests of model fit find significant improvement moving from Model 1 to Model 2 [ $\chi^2$  (16,  $N=2199$ )=87.15;  $p<.001$ ] as well as moving from Model 2 to Model 3 [ $\chi^2$ (3,  $N=2198$ )=106.75;  $p<.001$ ]. We then analyze the final model in racial subsets in Table 4. The logistic regression models use unweighted data, exclude respondents who failed the manipulation checks, and use listwise deletion for missing data.

To generate estimates representative of the U.S. population, we also conducted analyses with weighted data, without exclusions ( $N=2672$ ). Here we use the Stata svy suite of commands for weighted logit estimation and estimated the proportions shown in Figure 1. Combined, these analyses can both speak to our specific interest in diverse populations as well as reflect larger trends in the U.S. population.

## RESULTS

We found a substantial resistance to artificial intelligence. With weighting representative to the U.S. population, most respondents (52.9%) chose the human doctor and 47.1% chose AI clinic, with some variation along race and ethnicity, as shown in Figure 1.

### *Effects of Experimental Manipulations*

When AI was proven to be more accurate, respondents were substantially more likely to select it; this was one of the most pronounced effects (OR=1.48, CI 1.24-1.77,  $p<.001$ ). This effect arises from the manipulation that “Your doctor tells you that, based on scientific studies in leading journals, the AI system is proven more accurate at diagnoses compared to even specialist human physicians.”

When the PCP nudged the patient toward AI as the established option, patients were more likely to choose it (OR=1.25, CI: 1.05-1.50,  $p=.013$ ). This effect arises from the PCP saying that “For some time, I have been recommending the Med-X AI clinic for all my patients, and the nurse has already confirmed available appointments for you.”

The results also show greater AI uptake when the AI system is personalized to listen to the patient (OR=1.27, CI: 1.07-1.52,  $p=.008$ ). This effect is caused by the text manipulation: “The Med-X clinic staff will carefully listen to understand your lifestyle, preferences, values, and goals. At the Med-X clinic, you will have an extensive 45-minutes interview with a trained counselor, who will ask a range of questions to get your perspective on your healthcare.”

Other experimental conditions did not have significant effects, such as illness severity, whether the doctor would merely defer to the results or incorporate them into his or her opinion, whether the AI system would be racially or financially unbiased, and whether the AI used multiple measurements to tailor the treatment plan (all  $p > .05$ ).

\*\* Figure 1 About Here \*\*

\*\* Table 3 About Here \*\*

\*\* Figure 2 About Here \*\*

### ***Demographic and Attitudinal Associations***

Several demographic and attitudinal variables were also associated with AI uptake. Older respondents had significantly lower odds of choosing AI (OR: .99, CI: .987-.999,  $p=.03$ ). Conservatives have 35% less than equal odds than non-conservatives of choosing AI (OR: .65, CI: .52-.81,  $p<.001$ ) and for each unit increase in education, the odds are 1.10 greater for selecting an AI provider (OR: 1.10, CI: 1.03-1.18,  $p=.004$ ). Respondents who viewed religion as important had significantly lower odds of choosing AI (OR: .64, CI: .52-.77,  $p<.001$ ). In terms of the attitudinal variables, trust in AI companies was associated with significantly increased odds of choosing an AI provider (OR: 1.29, CI: 1.23-1.36,  $p<.001$ ).

### ***Associations with Race/Ethnicity***

As shown in Table 3 Model 3 and Figure 2, we found significant associations between AI uptake and respondent's race or ethnicity; specifically Black respondents had lower odds of selecting AI (OR=.73, CI: .55-.96,  $p=.023$ ) and Native Americans had higher odds of selecting AI (OR: 1.37, CI: 1.01-1.87,  $p=.041$ ) than White respondents.

Table 4 investigates the effects of other variables within race/ethnicity subsets. AI accuracy continued to be significant and positive regarding AI provider selection for Whites, Hispanics, and Asians. The PCP nudge towards AI as an established and easy option resulted in significantly greater odds of choosing AI for Hispanics and Asians, whereas learning that AI was unbiased by financial relationships with drugmakers and insurance companies increased AI selection for Native Americans. Interestingly, the racial subsets analyses revealed other demographic associations. Being conservative had the greatest effects on lowering AI selection for Whites and Asians, while the importance of religion significantly lowered odds for Whites and Blacks. For example, Black respondents who viewed religion as important have 57% less than equal odds of choosing the AI clinic, compared to those placing less importance on religion (OR: .43, CI: .27-.69,  $p=.001$ ).

**\*\*Table 4 About Here\*\***

## **DISCUSSION**

### ***Strengths and Limitations***

Our randomized, blinded experiment allows causal inference about the impact of the manipulations, and the factorial design allows strong statistical power, because every respondent provides an observation for every variable. Our diverse population from a high-quality survey sample allows extrapolation to the U.S. population, when weighted.

Prior work has shown that vignette experiments can have external validity (31, 32) but it may be harder for respondents to predict their own decisions in peculiar situations they have not experienced and where emotional salience is high. We speculate that in a real-world context, patients may be even more susceptible to their PCP's influence.

We tested a range of manipulations, asserting qualities of a hypothetical AI system ("Med-X"), such as its lack of racial bias or its proven accuracy beyond that of human specialists. However, these traits may or may not be true about any particular AI system.

While we tested two distinct clinical vignettes (leukemia and sleep apnea) with very different levels of severity and found similar results, other clinical situations may be different. Neither of these scenarios were presented as being acute or emergent, where patients may have greater deference to their PCP such situations (33, 34, 35). Other manipulations or other implementations thereof may also have different effects.

Finally, in addition to the exclusions noted above, we had pre-registered a plan to also exclude respondents who were familiar with the condition (leukemia or sleep apnea, depending on experimental assignment) because of their work in a healthcare setting, or because they, or a close family or friend, has had it. However, a larger than expected portion of the sample turned out to be familiar with the conditions (N=998) and these respondents were more familiar with sleep apnea (N=665) than leukemia (N=333). Excluding those respondents would unbalance the experimental conditions, and dramatically reduce statistical power, especially for racial/ethnic subgroups. Instead, we control for illness familiarity in the main models and include models that exclude illness familiarity in the supplemental materials. Results for experimental manipulations were similar, with AI accuracy, PCP nudging as established care, and AI listening to patients all significant predictors of uptake.

Error! Bookmark not defined.

### ***Findings and Implications***

Consistent with other studies, we found substantial resistance to the use of AI, what some have called “algorithm aversion” (5) Error! Bookmark not defined. or “robophobia” (4). Error! Bookmark not defined. Our study contributes to this literature by showing the robustness of this resistance to AI in a diverse population of patients.

We find this aversion to AI diagnosis across two levels of disease severity, within specific racial and ethnic groups, and even where the AI was proven to be more accurate and the PCP nudged the patient towards AI, though these efforts were significantly helpful. Some of our hypothesized attempts to mitigate this resistance were ineffectual. Most surprising, having the PCP emphasize that he or she would explain and incorporate the AI system’s advice into the ultimate treatment decision did not increase uptake of AI. In a world where AI is (or will have the potential to be) actually more accurate, our findings suggest that patients may suffer additional mortality and morbidity, and the healthcare system may suffer inefficiencies, due to patient resistance. The passage of time and expanded use of AI in a range of settings familiar to laypersons (such as self-driving cars) may help patients become more familiar with and supportive of AI in healthcare.

Respondents were asked to assume insurance coverage and out-of-pocket costs would be the same for either clinic. Insurers or healthcare systems may someday use these economic dimensions to direct or undermine patient choice of provider. Additionally, physicians or healthcare systems may deploy AI without giving the patient a choice or asking permission, especially where AI has become the standard of care. While such economic, professional, and managerial gatekeeping are common in a range of settings, they raise distinct ethical and legal concerns (36, 37), recognizing that patient autonomy and patient welfare may sometimes come into tension. In addition to our focus on patient perspectives, additional research is required on drivers of physician uptake of AI.

Our findings will be useful for the development of theory-based and evidence-driven recommendations for how physicians and patients might integrate AI into the informed consent process and into real world use and delivery of care. Patient resistance to AI diagnosis may impinge uptake in ways that

undermine treatment goals, but human physicians can support adoption, where the technology is designed with the patient experience in mind and supported by evidence of accuracy.

## REFERENCES

1. Savage N. How AI is improving cancer diagnostics. *Nature* [Internet]. 2019 March [cited 2021 May 10];2(94):1-10. Available from: <https://www.nature.com/articles/d41586-020-00847-2>
2. Callahan A, Fries JA, Ré C, Huddleston III JI, Giori NJ, Delp S, et al. Medical device surveillance with electronic health records. *NP J Digit Med* [Internet]. 2019 Sept. [cited 2021 May 10];2(94):1-10. Available from: <https://www.nature.com/articles/s41746-019-0168-z.pdf>
3. Xu J, Yang P, Xue S, Sharma B, Sanchez-Martin M, Wang F, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges, and future perspectives. *Hum Genet*. 2019 Feb;138(2);109-24.
4. Woods AK. Robophobia. *Univ Colo Law Rev*. 2021;93:1.
5. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J Exp Psychol*. 2015;144(1):114-26.
6. Meehl P. *Clinical vs. statistical prediction: a theoretical analysis and a review of the evidence*. Brattleboro, VT: Echo Point Books and Media; 1954.
7. Promberger M, Baron J. Do patients trust computers?. *J Behav Decis Making*. 2006 Nov;19:455-68.
8. Arkes H, Shaffer V, Medow M. Patients derogate physicians who use a computer-assisted diagnostic aid. *Med Decis Making*. 2007;27:189-202.
9. Shaffer V, Probst CA, Merkle EC. 2012. Why do patients derogate physicians who use a computer-based diagnostic support system?. *Med Decis Making*. 2012;33(1):108-18.
10. Longoni C, Bonezzi A, Morewedge C. Resistance to medical artificial intelligence. *J Consum Res*. 2019;46:629-50.
11. Longoni C, Bonezzi A, Morewedge C. Resistance to medical artificial intelligence is an attribute in a compensatory decision process: Response to Pezzo and Beckstead. *Judgm Decis Mak*. 2020 Jan;15:446-8.
12. Castelo N, Bos MW, Lehmann DR. Task-dependent algorithm aversion. *Journal of Marketing Research*. 2019 Oct;56(5):809-25.
13. Jonmarker O, Strand F, Brandberg Y, Lindholm P. The future of breast cancer screening: what do participants in a breast cancer screening program think about automation using artificial intelligence?. *Acta Radiol Open*. 2019 Dec;8(12):2058460119880315.
14. Palmisciano P, Jamjoom AA, Taylor D, Stoyanov D, Marcus HJ. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. *World Neurosurg*. 2020 Jun 1;138:e627-33.
15. Logg JM, Minson JA, Moore DA. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ Behav Decis Process*. 2019 Mar 1;151:90-103.

16. Mechanic D, Schlesinger M. The impact of managed care on patients' trust in medical care and their physicians. *JAMA*. 1996;275:1693–7.
17. Safran DG, Taira DA, Rogers WH, Kosinski M, Ware JE, Tarlov AR. Linking primary care performance to outcomes of care. *J Fam Pract*. 1998;47:213–20.
18. Gamble VN. A legacy of distrust: African Americans and medical research. *Am J Prev Med*. 1993 Nov-Dec;9(6 Suppl):35–38.
19. Fairchild AL, Bayer R. Uses and abuses of Tuskegee. *Science*. 1999 May 7;284(5416):919–921.
20. Shavers VL, Lynch CF, Burmeister LF. Racial differences in factors that influence the willingness to participate in medical research studies. *Ann Epidemiol*. 2002 May;12(4):248–256.
21. Street RL, O'Malley KJ, Cooper LA, Haidet P. Understanding concordance in patient-physician relationships: personal and ethnic dimensions of shared identity. *Ann Fam Med [Internet]*. 2008 [cited 2021 May 10];6(3):198–205. Available from: <https://doi.org/10.1370/afm.821>
22. Schoenthaler A, Montague E, Baier Manwell L, Brown R, Schwartz MD, Linzer M. Patient-physician racial/ethnic concordance and blood pressure control: the role of trust and medication adherence. *Ethn Health [Internet]*. 2014 [cited 2021 May 10];19(5):565–78. Available from: <https://doi.org/10.1080/13557858.2013.857764>
23. Peek ME, Gorawara-Bhat R, Quinn MT, Odoms-Young A, Wilson SC, Chin MH. (2013). Patient trust in physicians and shared decision-making among African-Americans with diabetes. *Health Commun [Internet]*. 2013 [cited 2021 May 10];28(6):616–23. Available from: <https://doi.org/10.1080.10410236.2012.710873>
24. Brax H, Fadlallah R, Al-Khaled L, Kahale LA, Nas H, El-Jardali F, Akl EA. Association between physicians' interaction with pharmaceutical companies and their clinical practices: A systematic review and meta-analysis. *PloS one*. 2017 Apr 13;12(4):e0175493.
25. Dehon E, Weiss N, Jones J, Faulconer W, Hinton E, Sterling S. A systematic review of the impact of physician implicit racial bias on clinical decision making. *Acad Emerg Med*. 2017;24(8):895–904.
26. Ntoutsis E, Fafalios P, Gadiraju U, Iosifidis V, Nejdil W, Vidal ME, et al. Bias in data-driven artificial intelligence systems—an introductory survey. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2020 Feb 3;10(3):e1356.
27. Pierson E, Cutler DM, Leskovec J. et al. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat Med [internet]*. 2021 Jan [cited 2021 May 10];27(1):136–40. Available from: <https://doi.org/10.1038/s41591-020-01192-7>
28. Noseworthy PA, Attia ZI, Brewer LC, Hayes SN, Yao X, Kapa S, Friedman PA, Lopez-Jimenez F. Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis. *Circ Arrhythm Electrophysiol [Internet]*. 2020 Mar [cited 2021 May 10];13(3):e007988. Available from: <https://doi.10.1161/CIRCEP.119.007988>

29. Sarwar S, Dent A, Faust K, Richer M, Djuric U, Van Ommeren R, Diamandis P. Physician perspectives on integration of artificial intelligence into diagnostic pathology. *NPJ Digit Med*. 2019 Apr 26;2(1):1-7.
30. Jussupow E, Spohrer K, Heinzl A, Gawlitza J. Augmenting medical diagnosis decisions? An investigation into physicians' decision making process with artificial intelligence. *Inf Syst Res*. 2020:tba.
31. Hainmueller J, Hangartner D, Yamamoto T. Validating vignette and conjoint survey experiments against real-world behavior. *Proc Natl Acad Sci U S A*. 2015 Feb 24;112(8):2395-400.
32. Ulrich CM, Ratcliffe SJ. Hypothetical vignettes in empirical bioethics research. In: Jacoby L, Siminoff LA, editors. *Empirical methods for bioethics: a primer*. Oxford, UK: Elsevier; 2008.
33. Boin A, Kuipers S, Overdijk W. Leadership in times of crisis: a framework for assessment. *International Rev Public Admin* [Internet]. 2013 [cited 2021 May 10];18(1):79–91. Available from: <https://doi.org/10.108012294659.2013.1080521>
34. Dirani KM, Abadi M, Alizadeh A, Barhate B, Garza RC, Gunasekara N, et al. Leadership competencies and the essential role of human resource development in times of crisis: a response to Covid-19 pandemic. *Human Resource Devel International* [Internet]. 2020 May 28 [cited 2021 May 10];23(4):1-15. Available from: <https://doi.org/10.1080/13678868.2020.1780078>
35. Schraagen JM, van de Ven J. Human factors aspects of ICT for crisis management. *Cogn Technol Work* [Internet]. 2011 [cited 2021 May 10];13(3):175-87. Available at: <https://doi.org/10.1007/s10111-011-0175-6>
36. Findley J, Woods AK, Robertson C, Slepian M. Keeping the patient at the center of machine learning in healthcare. *Am J Bioeth*. 2020 Nov 1;20(11):54-6.
37. Eyal N, Romain PL, Robertson C. Can rationing through inconvenience be ethical?. *Hastings Cent Rep*. 2018 Jan;48(1):10-22.

## TABLES

Table 1A – Presentation of Provider Choice

<b>Text Presented to Respondents</b>
<p data-bbox="300 384 750 411">“Your doctor offers either of two options:</p> <ul data-bbox="349 415 1325 703" style="list-style-type: none"><li data-bbox="349 415 1325 478">• You could go to the offices of Dr. Williams, [a hematologist-oncologist / NA], a specialist doctor who is trained to diagnose [leukemias / sleep disorders], or</li><li data-bbox="349 483 1325 703">• You could go to the office of Med-X, which is built around a proprietary computer system designed to diagnose [leukemias / sleep disorders]. Your blood and genetic information would be drawn by a nurse, and the medical analysis would be done entirely by a machine using artificial intelligence (AI). With every case it sees from tens of thousands of patients worldwide, the AI system gets more accurate in its diagnoses. If you visit the AI clinic, your data will be de-identified and then become part of the system.</li></ul> <p data-bbox="300 707 1325 798">[accuracy manipulation] Your insurance will only cover one of the services so you must choose which one to use to diagnose your possible [sleep apnea / leukemia]. Your out-of-pocket cost is the same either way. [other manipulations]”</p>

**Table 1B -- Experimental Conditions**

<b>Variable</b>	<b>Level 1</b>	<b>Level 2</b>
<b>Illness Severity</b>	Leukemia (“...There are several different types of leukemia. Some ... can spread to lymph nodes or the central nervous system if not treated. Leukemia can be fatal...”)	Sleep apnea (“...you are getting poor sleep at night, with very loud snoring, and sometimes it seems like you stop breathing or gasp for air during sleep....”)
<b>AI Accuracy</b>	No description.	“Your doctor tells you that, based on scientific studies in leading journals, the AI system is proven more accurate at diagnoses compared to even specialist human physicians.”
<b>AI Listens (Personal)</b>	No description	“The Med-X clinic staff will carefully listen to understand your lifestyle, preferences, values, and goals. ... you will have an extensive 45-minutes interview with a trained counselor, who will ask a range of questions to get your perspective on your healthcare.”
<b>AI Tailored (Personal)</b>	No description.	“The AI system’s advice will be tailored to you. [It] will incorporate 36 different measurements and attributes specific to you to generate a unique and personalized treatment plan, just for you.”
<b>AI Racial Unbiased</b>	No description.	“Although research suggests that human physicians can be biased by racial and ethnic stereotypes, the AI system has been carefully designed and tested to ensure that treatment recommendations are unbiased.”
<b>AI Financial Unbiased</b>	No description.	“Although research suggests that human physicians can be biased by their financial relationships with drugmakers and insurance companies, the AI system has been carefully designed and tested to ensure that treatment recommendations are unbiased.”
<b>PCP Incorporates AI Advice</b>	Defer-Portal: “Your doctor explains, ‘From either clinic, you will receive the results as an electronic message in your patient portal, the next day. It will tell you the diagnosis and what to do next.’”	Incorporate-Explain: “Your doctor explains, ‘When we get the results back from either clinic, I will explain them and talk them through with you. I will incorporate the results into my ultimate opinion on what we should do next.’”
<b>PCP Nudges toward AI</b>	None: “Your doctor says, ‘We can get you into either Dr. Williams or the Med-X AI clinic; it is your choice.’”	Default-Easy: “Your doctor says, ‘For some time, I have been recommending the Med-X AI clinic for all my patients, and the nurse has already confirmed available appointments for you. But if you prefer to see Dr. Williams, I can give you a referral instead. It is your choice.’”

**Note:** For the unbiased and personalization variables, when both types (race and financial, or listens and tailored) were presented, the text was concisely integrated into a single statement, as shown in the supplemental materials.

**Table 2. Descriptive Traits of Participants (N=2472)**

<u>Factor</u>	<u>Mean (SD) or N (%)</u>
Age	48.07 (17.18)
Female	1380 (55.83%)
Married	1183 (47.86%)
Employed Full-Time	853 (34.51%)
Less than 50K Income	1194 (54.27%)
High School or Less	864 (34.95%)
Some College or Associate Degree	777 (31.43%)
Bachelor’s Degree or More	831 (33.61%)
Conservative*	575 (23.27%)
Religion Important*	1536 (62.14%)
White	617 (24.96%)
Black	528 (21.36%)
Hispanic	489 (19.78%)
Asian	511 (20.67%)
Native American	327 (13.23%)
Good Health*	1897 (76.74%)
Have Provider*	2001 (80.95%)
Trust Provider*	1956 (79.13%)
Trust Hospital*	1735 (70.21%)
Trust AI Companies*	1346 (54.45%)

**Note:** Factors shown with asterisk (\*) are based on Likert-scales with top two values scored here, for example, good health (1=excellent/very good) and religion viewed as important (1=very/somewhat important). Income was measured on 16-point scale from less than \$10,000 to more than \$500,000), education was measured on a 6-point scale from less than high school to graduate school).

**Table 3. Logistic Regression Predicting AI Provider Choice Versus Human Physician**

	<b>Model 1: Experimental Conditions</b>		<b>Model 2: + Demographics</b>		<b>Model 3: + Trust</b>	
Illness Severity	0.96	(0.08)	1.00	(0.09)	0.98	(0.09)
Accuracy	1.51***	(0.12)	1.54***	(0.14)	1.48***	(0.13)
Incorporation	0.92	(0.07)	0.91	(0.08)	0.88	(0.08)
Established	1.28**	(0.10)	1.28**	(0.11)	1.25*	(0.11)
Tailored	1.06	(0.09)	1.07	(0.09)	1.11	(0.10)
Listens	1.20*	(0.10)	1.23*	(0.11)	1.27**	(0.12)
Race Unbiased	1.08	(0.09)	1.07	(0.09)	1.10	(0.10)
Financial Unbiased	1.13	(0.09)	1.12	(0.10)	1.10	(0.10)
Order	1.20*	(0.10)	1.20*	(0.11)	1.13	(0.10)
Illness Experience	1.12	(0.10)	1.18 <sup>+</sup>	(0.11)	1.17	(0.12)
Black			0.84	(0.11)	0.73*	(0.10)
Hispanic			1.00	(0.14)	0.88	(0.12)
Asian			0.94	(0.13)	0.85	(0.12)
Native			1.26	(0.19)	1.37*	(0.21)
Income			1.02	(0.02)	1.02	(0.02)
Age			0.99*	(0.00)	0.99*	(0.00)
Conservative			0.61***	(0.07)	0.65***	(0.08)
Religion Important			0.67***	(0.06)	0.64***	(0.06)
Sex			1.00	(0.09)	0.98	(0.09)
Education			1.07*	(0.04)	1.10**	(0.04)
Married			0.96	(0.09)	0.97	(0.10)
Employed Full-Time			1.03	(0.10)	1.05	(0.11)
Health			1.09	(0.12)	0.99	(0.11)
Have Provider			1.12	(0.13)	1.08	(0.14)
Trust Provider					0.96	(0.03)
Trust Hospital					0.95+	(0.03)
Trust AI Companies					1.299***	(0.03)
N	2471		2199		2198	

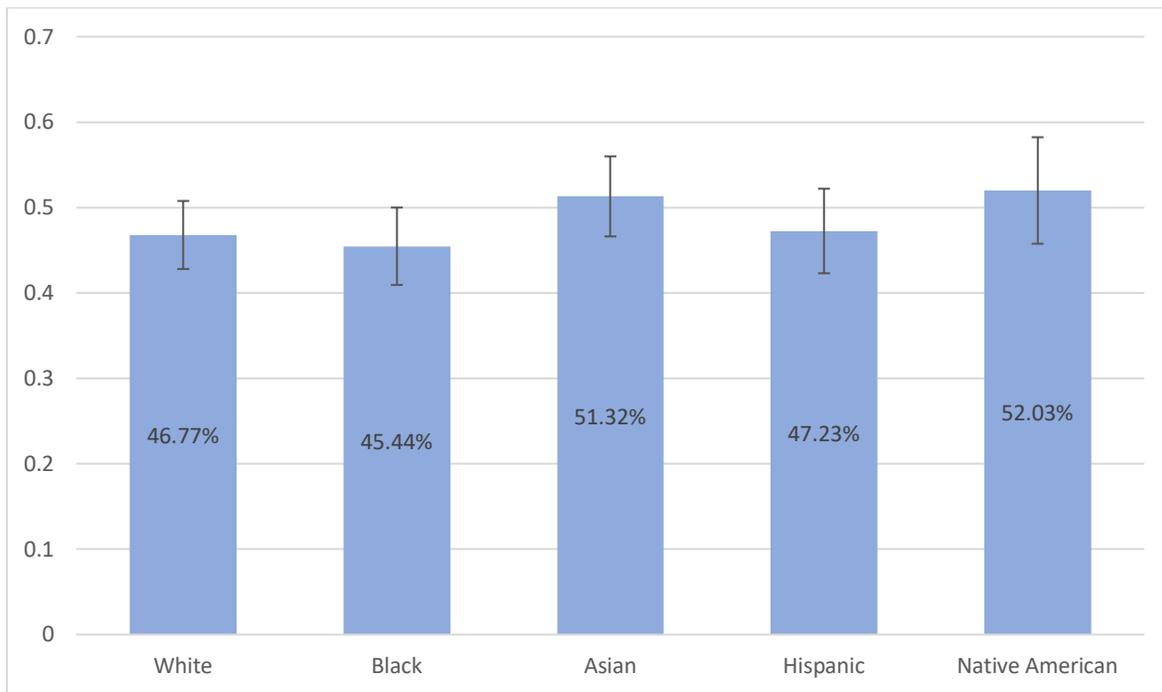
Notes: Odds Ratios; Standard errors in parentheses <sup>+</sup>  $p < .10$ , \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . For race, Whites are the reference group. Income (16-point scale from less than \$10,000 to more than \$500,000), age (years), education (6-point scale from less than high school to graduate school), sex (1=male, 0=female). Variables are dichotomized on top two values, e.g., good health (1=excellent/very good, 0=fair/poor) and religion viewed as important (1=very/somewhat important, 0=not too/not at all important). Analysis excludes respondents that failed an attention check about whether respondent could correctly recall the disease presented in the vignette, and those reporting that they did not understand the vignette (bottom two levels on 10-point scale).

**Table 4. Logistic Regression in Respondent Race/Ethnicity Subsets Predicting AI Provider Choice Versus Human Physician**

	<b>White</b>		<b>Black</b>		<b>Hispanic</b>		<b>Asian</b>		<b>Native</b>	
Illness Severity	0.97	(0.20)	1.05	(0.23)	1.01	(0.22)	0.78	(0.17)	0.97	(0.27)
Accuracy	1.76**	(0.33)	1.42 <sup>+</sup>	(0.28)	1.68*	(0.36)	1.69*	(0.37)	0.82	(0.22)
Incorporation	0.97	(0.18)	0.86	(0.17)	0.81	(0.17)	0.85	(0.18)	0.70	(0.18)
Established	1.16	(0.22)	0.91	(0.18)	1.57*	(0.33)	1.63*	(0.35)	1.15	(0.30)
Tailored	1.11	(0.21)	1.33	(0.27)	1.09	(0.23)	1.14	(0.25)	0.92	(0.24)
Listens	1.25	(0.24)	1.45 <sup>+</sup>	(0.29)	1.22	(0.25)	1.10	(0.23)	1.48	(0.40)
Race Unbiased	1.10	(0.21)	1.17	(0.23)	1.49 <sup>+</sup>	(0.32)	0.85	(0.18)	0.95	(0.25)
Financial Unbiased	0.90	(0.17)	1.11	(0.23)	1.15	(0.24)	1.13	(0.24)	1.86*	(0.50)
Income	1.00	(0.03)	1.02	(0.04)	1.07 <sup>+</sup>	(0.04)	1.03	(0.03)	0.96	(0.05)
Age	1.00	(0.01)	0.99	(0.01)	0.99	(0.01)	0.99	(0.01)	0.99	(0.01)
Conservative	0.58*	(0.14)	0.82	(0.25)	0.69	(0.18)	0.55*	(0.16)	0.60 <sup>+</sup>	(0.18)
Religion Important	0.59**	(0.12)	0.43***	(0.10)	0.79	(0.19)	0.66 <sup>+</sup>	(0.15)	0.80	(0.24)
Sex	0.81	(0.16)	1.43 <sup>+</sup>	(0.30)	1.44 <sup>+</sup>	(0.32)	0.68 <sup>+</sup>	(0.15)	0.75	(0.21)
Education	1.10	(0.08)	1.13 <sup>+</sup>	(0.08)	1.15 <sup>+</sup>	(0.09)	1.03	(0.09)	1.31*	(0.15)
Married	1.00	(0.21)	0.74	(0.17)	1.14	(0.25)	0.75	(0.18)	1.48	(0.42)
Employed Fulltime	1.02	(0.22)	1.13	(0.26)	1.07	(0.25)	1.41	(0.34)	0.62	(0.21)
Health	1.27	(0.30)	1.07	(0.26)	0.55*	(0.15)	1.04	(0.30)	1.13	(0.33)
Have Provider	1.01	(0.30)	1.65 <sup>+</sup>	(0.46)	0.76	(0.21)	1.01	(0.30)	1.54	(0.64)
Trust Provider	0.94	(0.06)	0.94	(0.06)	0.94	(0.06)	1.02	(0.07)	0.97	(0.07)
Trust Hospital	0.92	(0.05)	0.99	(0.07)	0.98	(0.06)	0.83*	(0.06)	0.99	(0.07)
Trust AI Companies	1.36***	(0.08)	1.20**	(0.07)	1.36***	(0.08)	1.36***	(0.09)	1.34***	(0.09)
Order	1.05	(0.20)	1.17	(0.23)	1.01	(0.21)	1.50 <sup>+</sup>	(0.32)	0.99	(0.27)
Illness Experience	1.12	(0.23)	1.41	(0.32)	0.74	(0.17)	1.29	(0.30)	1.28	(0.38)
<i>N</i>	542		479		442		437		298	

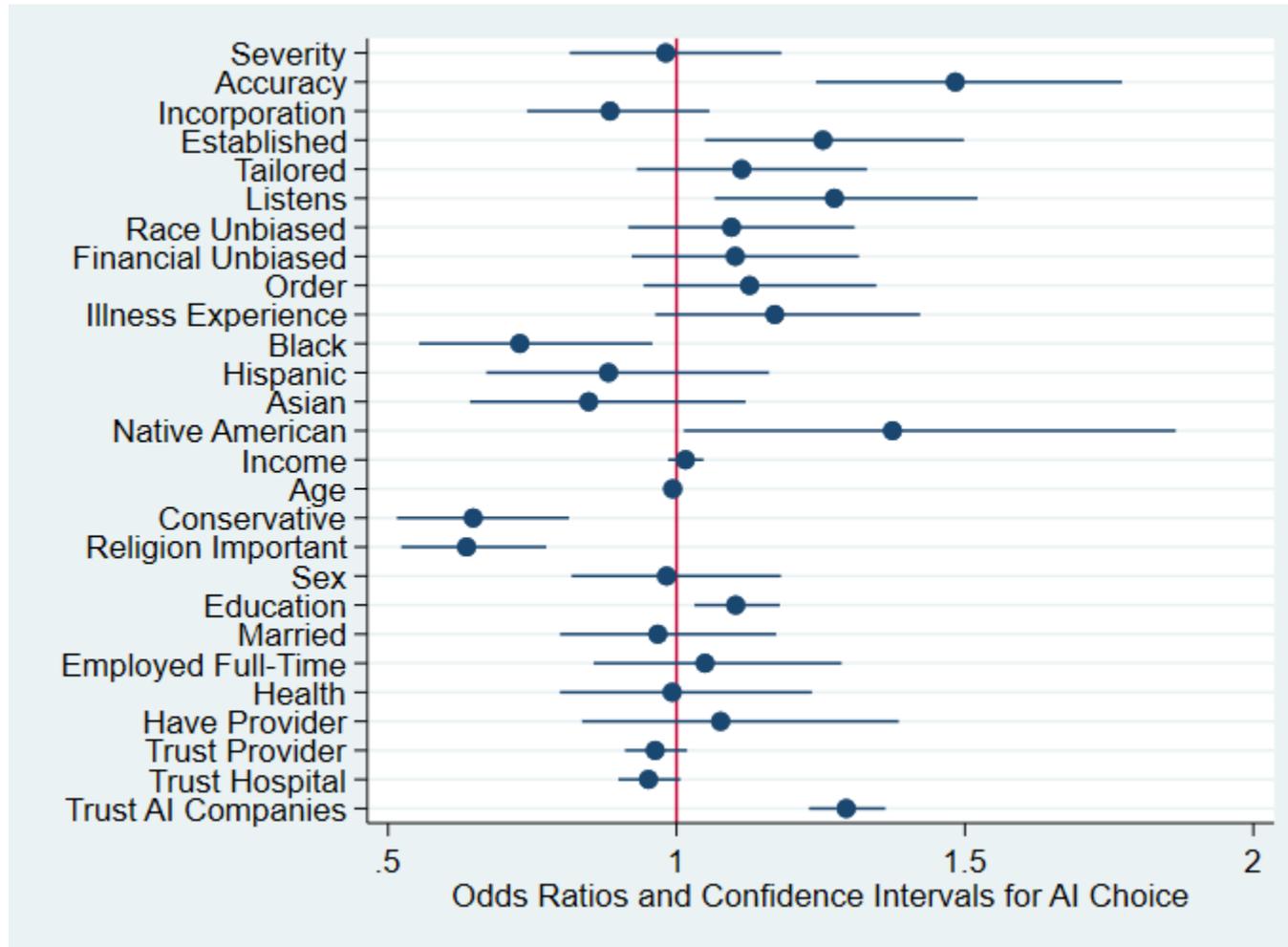
Notes: Odds Ratios; Standard errors in parentheses <sup>+</sup>  $p < .10$ , \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . Income (16-point scale from less than \$10,000 to more than \$500,000), age (years), education (6-point scale from less than high school to graduate school), sex (1=male, 0=female). Variables are dichotomized on top two values, e.g., good health (1=excellent/very good, 0=fair/poor) and religion viewed as important (1=very/somewhat important, 0=not too/not at all important).

**Figure 1. Proportion of Respondents Selecting an AI Provider with 95% Confidence Intervals by Respondent Race / Ethnicity (N=2672, Weighted to Represent U.S. Population)**



**Notes:** Using the 2018 American Community Survey as the reference, the sample was weighted age, gender, race/ethnicity, years of education, and region to produce estimates for the U.S. population

**Figure 2. Odds of Selecting an AI Provider by Experimental Condition, Demographics, and Attitudinal Variables, With 95% Confidence Intervals (N=2198)**



*Notes:* Analysis excludes respondents that failed an attention check requiring recall of the disease presented in the vignette, and those reporting that they did not understand the vignette (bottom two levels on 10-point scale). See Table 3, Model 3 for details.

