https://doi.org/10.1038/s44184-025-00124-y

Assessing the impact of AI on physician decision-making for mental health treatment in primary care

Check for updates

Katie Ryan, Hyun-Joon Yang, Bohye Kim & Jane Paik Kim ✉

AI models may soon be poised to recommend mental health treatments or referrals in primary care, yet little is known regarding their impact on physician decision-making. In this web-based study, primary care physicians ($n = 420$) were presented with a clinical scenario describing a patient with psychiatric symptoms, an AI tool for referring or prescribing, and the recommendation of the AI. A sequentially randomized vignette method was used to test the impact of initial assessments and AI output on physician decision-making patterns. Physicians were significantly more likely to change their decisions when the AI recommendation was misaligned with their initial assessment, especially when AI recommended treatment. There was no difference between the change-in-decision rate of physicians who received an AI recommendation to not treat, indicating that the direction of AI recommendations may influence physician decision-making, and raising important considerations for how physician decisions may be anticipated in the context of AI.

Advances in AI present the opportunity to enable front-line, primary care physicians to more efficiently and accurately diagnose, refer, and treat patients who present with specialized conditions[1]. While still at an early stage, clinical decision support systems employing AI (AI-CDSS) have demonstrated how AI's integration into primary care settings can result in earlier and more accurate diagnoses, lessening the burden on specialists and allowing patients earlier access to needed care[2–5]. The integration of AI-CDSS into primary care is especially promising for the field of mental health, where shortages of providers have led to primary care physicians being the de-facto provider for mental health screening and diagnosing services, and creating, as described by Lee et al.[6], "an urgent need for AI to help identify high-risk individuals and provide interventions to prevent and treat mental illnesses"[6–8].

Using AI for the prediction of anxiety, depression, and mental health crises has been explored, yet little is understood about how the integration of similar systems into the primary care setting may influence providers' decision making and resultant patient outcomes[9–12]. Studies have identified trust, explainability, and understanding as factors which impact physicians' stated willingness to use AI-CDSS and incorporate it into their decision-making processes, but few have in fact empirically tested whether and how the integration of AI-CDSS influences physician decision making[13,14]. Studies that have tested these questions in real and hypothetical settings have found that, while correct AI predictions tend to improve physicians' diagnostic accuracy, incorrect AI predictions are often not dismissed and can decrease physician accuracy below what it would have been without the use of AI[15–18]. While earlier studies found that clinicians often fail to adopt AI recommendations, more recent experiments found the opposite tendency, especially amongst physicians with less task expertise[15,19]. This heterogeneity of findings in the literature leaves this question still in need of further examination.

Moreover, a major potential benefit of AI-CDSS for primary care is the expansion of specific task expertise amongst physicians who typically provide more generalized care. As noted in the study by Gaube et al.[15], it is precisely this group of physicians who may be the most likely to incorporate AI recommendations into their own decision making given the volume of different medical and psychiatric conditions they encounter in practice. The anticipation of primary care physicians' decision-making patterns therefore represents a notable gap in knowledge which must be addressed for the safe, effective, and ethical deployment of AI in mental health contexts[20,21].

Given the high stakes nature of experimenting or observing in the wild physician decision making with patients, hypothetical scenarios offer a lower risk in exposing physicians to such scenarios. In this study, we examined primary care physician decision-making in hypothetical clinical scenarios involving AI-CDSS for mental health to understand how their clinical decision-making may be influenced by AI, and if the initial assessment of physicians and the output of the AI-CDSS impacted these decisions. For the current paper, our focus was on clinical decision support systems employing deep learning algorithms in AI for prediction, though the capabilities of AI in the context of clinical decision support may be expected to be much broader. To address some of the limitations that are typically

Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA. ✉e-mail: janepkim@stanford.edu

present in hypothetical vignette studies, we applied a novel study design utilizing an adaptation of the sequential multiple assignment randomization trial (SMART) design[22]. Through an exploratory analysis of physicians' decision-making patterns with AI-CDSS, we identified outcomes relating to the change between physicians' initial clinical and final assessments (pre- and post-AI) and the self-reported influence of AI on their final decision, and their consistency amongst these factors.

Results

Experiment

Participants in this web-based study were physicians in the United States with a subspecialty in either family medicine or internal medicine ($n = 420$). Participants were presented with a SMART vignette describing a hypothetical mental health clinical scenario involving a patient with symptoms of ADHD or depression and an AI-CDSS intended to recommend a decision to the physician[22,23]. The SMART vignette featured three time points which randomized participants to hypothetical scenarios ("subvignettes") that were varied by: (1) clinical decision type (ADHD referral vs. SSRI prescription); (2) amount of information about the AI-CDSS (short explanation vs. long explanation); and (3) type of AI recommendation (does not recommend treatment vs. recommends treatment) (See Fig. 1). A question set corresponding to each scenario was asked at each randomization time point. The complete text of the subvignettes and corresponding questions are included in the Supplementary File.

Outcomes. In this paper, we focused on the following outcomes: (1) changes in clinical decisions (time 1 vs. time 3), (2) physicians' self-report regarding whether they would have made a different decision without the involvement of AI. As an exploratory outcome, we developed a new outcome of "fidelity" defined by the alignment between the physicians' initial assessment and their final self-reported decision had AI not been involved.

Aims. We examined how these decision-making outcomes varied by AI recommendation and the initial physician assessment. Specifically, we assessed whether 1) the change in clinical decisions from the initial to final time point is explained by the initial assessment and the type of AI recommendation; 2) the response to whether the physician would have made a different decision without AI's involvement is impacted by the aforementioned factors. As an exploratory aim, we assessed how "fidelity" between this decision without AI and initial assessment is affected by the amount of information on the AI, AI recommendation, and physicians' familiarity with AI.

Change in decision from initial assessment to final decision

Overall, physicians who received an AI recommendation that did not align with their initial assessment were substantially more likely to change their decisions. Nearly two thirds of physicians (66.7%) who received an AI recommendation that misaligned with their initial assessment of not treating the patient ended up changing their decision to treat the patient. Moreover, physicians who initially assessed the hypothetical patient as an unsuitable candidate for an SSRI prescription (or ADHD referral) and yet later received an AI recommendation to prescribe/refer treatment were 9-fold more likely to change their decision than those who received an AI recommendation in alignment with their initial judgment to do nothing (95% CI, 2.805–28.873%). Conversely, physicians who assessed the hypothetical patient as a good candidate, and yet later received an AI recommendation to not treat were 5.72 times more likely to change their decision than those who received an AI recommendation to prescribe (95% CI, 0.078–0.383). Approximately

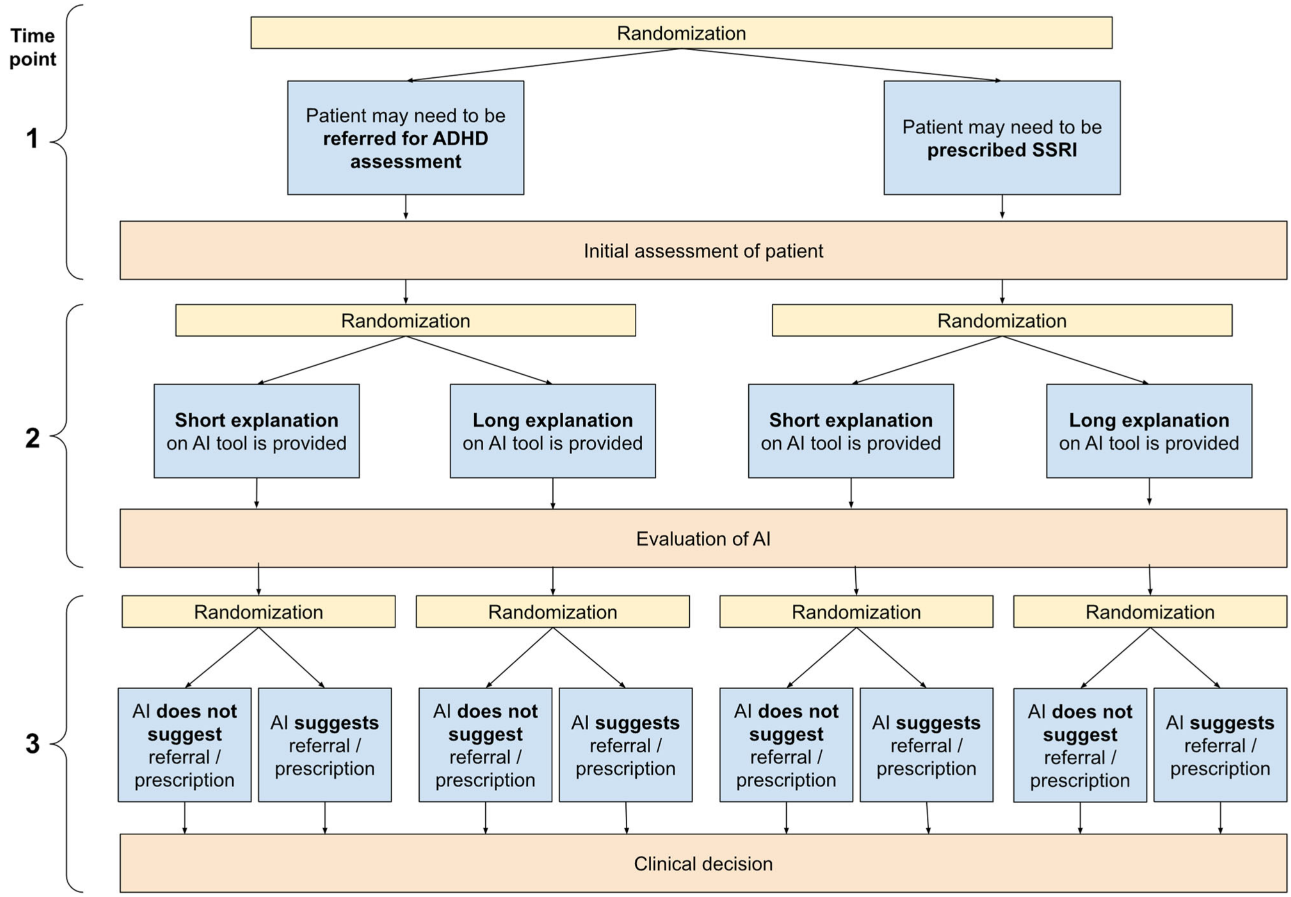
Time point
1
2
3
Randomization
Patient may need to be **referred for ADHD assessment**
Patient may need to be **prescribed SSRI**
Initial assessment of patient
Randomization
Randomization
Short explanation on AI tool is provided
Long explanation on AI tool is provided
Short explanation on AI tool is provided
Long explanation on AI tool is provided
Evaluation of AI
Randomization
Randomization
Randomization
Randomization
AI **does not suggest** referral / prescription
AI **suggests** referral / prescription
AI **does not suggest** referral / prescription
AI **suggests** referral / prescription
AI **does not suggest** referral / prescription
AI **suggests** referral / prescription
AI **does not suggest** referral / prescription
AI **suggests** referral / prescription
Clinical decision

Fig. 1 | Setup of the SMART Survey

21.55% of physicians who assessed the patient as a good candidate and received an AI recommendation that misaligned with their initial assessment ultimately changed their decision to do nothing.

The initial physician assessment did not have an impact on the odds of physicians changing their decisions when the AI recommendation was “no treatment” (95% CI, 0.477–3.205). Physicians who received a recommendation to treat were 98% less likely to change their decisions if their initial assessment was aligned with this recommendation (95% CI, 0.008–0.067). Table 1 reports the conditional odds ratios of changing their decision, depending on the physicians’ initial assessment and type of AI recommendation. Figure 2 shows the percentage of physicians who changed their opinion based on their initial assessment and the AI recommendation.

Table 1 | Conditional odds ratios of change in decision based on initial assessment and certainty of AI output

Outcome: Change in Decision from Initial Assessment to Final Decision	AI Recommendation Odds (%)		Odds Ratio [Confidence Interval]
Physician Assessment Odds (%)	Do not refer/ prescribe	Refer/ prescribe	
Not a Good Candidate	0.222 (18.18%)	2.000 (66.67%)	9.000 [2.805, 28.873]
Good Candidate	0.275 (21.55%)	0.048 (4.55%)	0.173 [0.078, 0.383]
Odds Ratio [Confidence Interval]	1.236 [0.477, 3.205]	0.024 [0.008, 0.067]	-

Significant odds ratios are underlined with the 95% confidence interval in brackets. Confidence intervals that do not include 1 indicate statistical significance.

Self-reported final decision without AI involvement

A similar though less drastic trend was found for self-reported final decisions in that physicians who received an AI recommendation that misaligned with their initial assessment were more likely to report that their final decision would have been different without AI. Among physicians who received a recommendation that misaligned with their initial judgment to not treat, approximately 26.7% reported that their decision would have been different had AI not been involved. Moreover, physicians who initially assessed the hypothetical patient as an unsuitable candidate for an SSRI prescription (or ADHD referral) and yet later received an AI recommendation to prescribe/refer treatment were 5.63 fold more likely to self-report a different decision than those who received an AI recommendation in alignment with their initial judgment to do nothing (95% CI, 1.090, 29.144). Conversely, physicians who assessed the hypothetical patient as a good candidate, and yet later received an AI recommendation to not treat were 4.94 times more likely to self-report a different decision than those who received an AI recommendation to prescribe (95% CI, 0.103, 0.397).

The initial physician assessment had a significant impact on the odds of reporting a different decision without AI with both types of recommendations from AI. Physicians who received a recommendation to treat were remarkably (i.e. 80%) less likely to report they would have made a different decision without AI if their initial assessment was aligned with this recommendation (95% CI, 0.074–0.547). Physicians who received a recommendation to not treat were 5.59 times more likely to report a different decision without AI if their initial assessment was misaligned with this recommendation (95% CI, 0.074–0.547). Table 2 reports the conditional odds ratios of whether the final decision would have been different without AI involvement, depending on the physicians’ initial assessment and type of AI recommendation. Figure 2 shows the percentage of physicians who self-reported they would have made a different decision without AI based on their initial assessment and the AI recommendation.

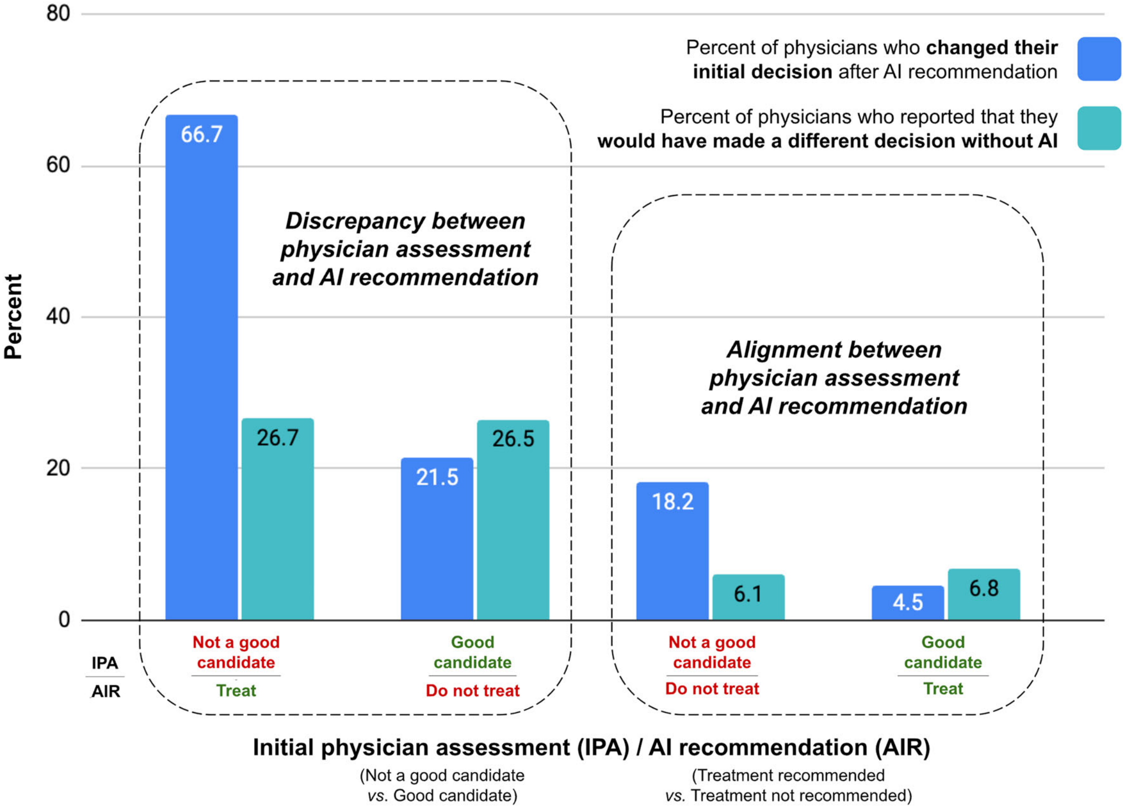

Fig. 2 | Percentage of physicians who changed decisions and physicians who self-reported they would have made different decisions without AI

Table 2 | Odds and odds ratios of outcome relating to self-reported final decision without AI involvement based on initial assessment and AI output

Outcome: I would have made a different decision if MedAssist was not involved	**AI Recommendation Odds (%)**		**Odds Ratio [Confidence Interval]**
Physician Assessment Odds (%)	Do not refer/ prescribe	Refer/ prescribe	
Not a Good Candidate	0.065 (6.10%)	0.364 (26.69%)	5.636 [1.090, 29.144]
Good Candidate	0.361 (26.53%)	0.073 (6.80%)	0.203 [0.103, 0.397]
Odds Ratio[Confidence Interval]	5.594 [1.289, 24.269]	0.201 [0.074, 0.547]	-

Significant odds ratios are underlined with the 95% confidence interval in brackets. Confidence intervals that do not include 1 indicate statistical significance.

Fidelity

Overall, physicians who received an AI recommendation aligned with their initial assessment were more likely to have fidelity in their decisions. For example, physicians who received an AI recommendation that was aligned with their assessment to do nothing were 10.56 fold more likely to observe fidelity than when receiving a misaligned recommendation. Stated alternatively, physicians who received an AI recommendation that misaligned with their initial assessment to not treat were 90.8% less likely to have a match between their initial and expected assessments (95% CI, 0.026, 0.329). Conversely, physicians who assessed the hypothetical patient as a good candidate, and later received an AI recommendation to treat were 2.90 times more likely to observe fidelity than those who received an AI recommendation to not prescribe (95% CI, 1.606, 5.25).

When physicians received an AI recommendation to treat the patient, the initial assessment had a remarkable impact on fidelity (OR 13.16; 95% CI, 5.471, 31.687). There was no impact of the initial assessment on fidelity when the AI recommendation was to not treat the patient (OR 0.417; 95% CI, 0.139, 1.250).

Discussion

Advancements in AI-CDSS offer the potential for primary care physicians to more accurately and efficiently diagnose, refer, and treat health conditions that may benefit from specialty input. These opportunities are especially promising for the field of mental health, where a dearth of specialists and barriers to access mental health care services has resulted in patients seeking mental health care solutions from primary care physicians[6,7]. The safe integration of AI and decision-support tools in mental healthcare, however, necessitates a more thorough understanding of how these technologies influence physician decision-making, and also patient outcomes. By employing the SMART vignette methodology described by Kim et al.[23], we studied how different factors—details about an AI-CDSS, its recommendation - impact physician decision making in a primary care setting without actually exposing patients at risk[22,23].

Physicians in this study demonstrated a notable difference in their likelihood to change their initial judgment regarding their decision to treat a hypothetical patient who presents with potential symptoms of ADHD or depression when accounting for the alignment between their initial clinical assessment and the subsequent AI recommendation. Physicians who were assigned an AI recommendation to further treat the patient were significantly more likely to change their opinion when there was a discrepancy between their initial clinical assessment and the AI recommendation, and significantly more likely to maintain their original opinion when there was alignment between their initial assessment and the AI recommendation. There was, however, no significant difference between the change-in-decision rate of physicians who received an AI recommendation to not treat, regardless or whether it was in alignment with their initial assessment. While not included in the results, a similar trend was observed when additionally accounting for type of clinical decision.

This finding is notable in that it suggests that, when presented with an AI recommendation that involves taking further action, physicians tend to react by either changing their initial opinion to match that of the AI, or by standing firm in their initial opinion which was supported by the AI. When presented with an AI recommendation that involved withholding further treatment, or not taking further action, however, physicians did not demonstrate a similar level of willingness to incorporate the AI advice into their clinical decision-making. This both supports and furthers Gaube et al.'s[15] finding that physicians with less task expertise are more likely to agree with advice provided by AI[15]. In the case of primary care physicians in our study, it appears that this tendency to agree with AI may be particularly true in cases where an AI recommends action.

The results of our analysis further suggest that, once action is recommended by AI, physicians may feel a compulsion to conform to it, regardless of their initial opinion. When action is not recommended, physicians may feel more open to being flexible in their judgment. These findings could be interpreted in the context of physician liability. Questions regarding the liability of physicians in an age of clinical AI have already been raised, with the increased vulnerability of physicians who go against an AI recommendation being specifically highlighted[24,25]. Interestingly, physicians in our study appeared to be fairly comfortable going against the AI recommendation when it recommended no action. Particularly given the relatively low-risk nature of the next steps in our clinical scenarios, it may not be surprising that this group of physicians were willing to prescribe/refer against the advice of the AI. However, the inverse was not true, and when the AI recommended action, physicians, regardless of their initial opinion, were remarkably likely to conform to the AI's recommendation. It may be that physicians feel an increased sense of liability when presented with a recommended action, and they may feel that they are making a riskier decision (both for the patient and their own liability) if they decline to pursue a course of action that is explicitly recommended to them.

An additional finding of this study was that physicians who received an AI recommendation that misaligned with their initial assessment were more likely to report that their final decision would have been different without AI. While this finding on its own is not surprising, the difference in percentage of physicians who initially assessed the patient as a poor candidate for further treatment but ultimately chose to treat the patient after receiving a recommendation from the AI to do so (66%), and the percentage of these physicians who agreed that the AI influenced their final decision (27%) is notable. This seems to indicate that, amongst the group of physicians who were most likely to change their decision, there was less awareness of, or less willingness to admit to, the AI's influence on their decision. We again consider physician liability as a potential explanation for this behavior. Physicians' awareness that AI output will be recorded in a patients' file along with the final clinical decision has had a demonstrated effect on the extent to which physicians incorporate and defer to the AI advice, even if it is incorrect[26]. It may be that physicians in our study whose initial assessment of not to treat was contradicted by the AI recommendation felt the need to "cover their tracks," whether intentionally or unintentionally, in order to provide a stronger justification for why their final decision did not align with their initial assessment.

Understanding the cognitive behaviors of physicians that may contribute to or explain these tendencies related to clinical decision is crucial and has previously been explored by Jussupow et al.[18]. Their study found that physicians whose initial beliefs are confirmed by AI tend to have greater confidence in their original diagnoses, whereas physicians whose initial assessments do not align with the AI recommendation tend to "make a final diagnostic decision based on their dominant belief" regarding the capability of the AI. While our study provides further support regarding the direction of these interactions, it additionally indicates that content of the AI recommendation (i.e., whether or not it recommends action) may be an additional factor which can sway a physician to incorporate or ignore AI advice in their clinical decision making.

Analyses in this study further revealed that the types of clinical scenarios (i.e., ADRD referral vs. SSRI prescription) did not have an impact on physician decision-making, nor did the amount of information provided about the AI-CDSS. There have been mixed results regarding how increased explainability and information sharing affect physician trust and accuracy when using AI-CDSS, with an increasing number of studies acknowledging that increased explainability may not necessarily improve physician accuracy or patient outcomes[27–30]. Our study, which was able to simultaneously test the effect of multiple factors on physician decision making, further indicates potential limits to the benefits of increased information sharing, and instead suggests that other factors relating to physician alignment with the AI result and the content of AI recommendation may be more relevant to how physicians make their final decisions, potentially overwhelming any effect that increased explainability or information sharing may have.

Limitations of this current study include the use of hypothetical vignettes and self-reported outcomes. The outcomes were not observed in real settings and thus may not generalize to real time scenarios. We additionally acknowledge that physician willingness to incorporate AI-CDSS into their decision-making is likely influenced by many other factors beyond what we tested in this study, namely: AI transparency; hospital or employer expectations, oversight, and policies; and liability and legal protections liability. Future research questions may wish to test these factors, as well as whether the trends we found in relatively common mental health related decision-making scenarios persist in situations of higher risk. It is not clear as to whether physicians utilizing AI in such situations would be more or less likely than the physicians in our study to defer to the AI if it presented a recommendation that contradicted their initial assessment.

Given the high stakes and potential consequences of widespread integration of AI-CDSS, we argue there is great value in anticipating decision making and understanding factors impacting decision changing among physicians prior to the deployment of AI-CDSS in clinical contexts. Filling this gap can inform the development of more effective AI tools that are better aligned with physician needs and concerns, ultimately leading to improved patient outcomes and more efficient healthcare delivery. Through the use of a novel survey method, our study identified the factors of physician alignment with AI, and the type of AI recommendation, as significant influences on physician decision making in a primary care setting which should be considered further as AI is increasingly incorporated into healthcare.

Methods

Participants

The American Medical Association (AMA) Physician Professional Data (formerly known as the Physician Masterfile) was used to identify active physicians to invite to participate in this web-based survey study[31]. A random sample of 10,000 physicians from the AMA Physician Professional Data file who met the inclusion criteria of being located in the United States and having a listed subspecialty of Family Medicine or Internal Medicine were identified. Between 5/16/2023 and 12/7/2023, up to four emails and three mailed letters containing a brief description of our study, a weblink to the online survey, and unique respondent codes were sent to the physicians identified as part of the random sample. We received 513 total survey responses, 420 of which were complete and deemed eligible for inclusion in the final data set after review. See Table 3 for full participant characteristics. The demographic makeup of this sample is largely representative of the population of U.S. physicians in Internal and Family medicine as documented in the 2022 Physician Specialty Data Report by the Association of American Medical Colleges.

This study received human subjects research ethics approval from the Stanford University Institutional Review Board. All participants provided electronic informed consent and were compensated for their time and effort with a $10 Amazon.com gift code at the time of survey completion.

SMART vignette study design

This study employed a SMART vignette design, which is an online vignette-based survey adaptation of the sequential multiple assignment randomization trial (SMART) design[22]. SMART vignettes, proposed by Kim and Yang[23], feature novel properties previously not applied in the traditional vignette survey setting, namely sequential randomization and adaptive allocation. In traditional vignette studies, the full vignette with a predetermined combination of characteristics and all questions are presented at once. However, in SMART vignettes, the vignettes are divided into subvignettes, with each subvignette reflecting a particular characteristic. The subvignettes are then shown sequentially along with their corresponding question set over discrete time points.

Sequential randomization is particularly suitable when "the order of outcomes carries significance, or when subsequent outcomes depend upon the previous outcome"[22]. While traditional vignettes allow inferences concerning the joint impact of the interventions (i.e. vignette characteristics) on the outcomes, SMART vignettes allow inferences on the impact of each intervention as well as previous outcomes on following outcomes. SMART vignettes also employ adaptive allocation, which leverages Efron's biased coin design to mitigate imbalance between groups often observed in traditional vignette studies. In fact, conventional designs use complete randomization to present only a sample of the full vignette population in order to reduce response burden. However, doing so often leads to imbalance. Adaptive allocation encourages randomizations to the underrepresented group using a prespecified probability. The advantage of this biased randomized strategy is that we can administer a sample of the full set of vignettes without sacrificing balance between groups.

The SMART vignette survey was administered using the platform developed by Kim et al. and hosted on "smartvignettes.stanford.edu." The vignette survey featured three randomization time points which randomized participants to subvignettes reflecting hypothetical scenarios which were varied by: (1) clinical decision type (prescription of SSRI or referral for ADHD assessment); (2) amount of information on the AI tool (short or long); and (3) type of recommendation (treat or not) (See Fig. 1). The probability of the biased coin was set to $p = 2/3$ for adaptive allocation. For clinical decision type, the two scenarios (i.e. prescription or referral) were selected for the reasons that they were both related to psychiatric concerns and because we hypothesized these decision types conferred a different level of impact on the patient and could possibly influence whether or not physicians incorporated recommendations made by AI-CDSS.

Measures

Upon accessing the survey and providing electronic informed consent, participants completed brief baseline questionnaires regarding their demographic information and professional experience, familiarity and professional experience with AI, and general attitudes toward the use of AI in medicine. At the completion of these questionnaires, participants were navigated to the start of the SMART vignette. A series of questions were presented to participants at each randomization time point within the SMART vignette. A primary question was chosen for each time point to be used for adaptive allocation. The complete subvignettes and their accompanying questions can be found in the Supplementary File.

Time point 1. At the first time point, physicians were presented with a subvignette which described a clinical scenario involving a 33-year-old female patient. Physicians were randomized to receive a description of a patient who presented with symptoms of either depression or ADHD. Physicians were asked 5 questions about their initial clinical assessment of the hypothetical patient and their confidence in this assessment given the patient's symptoms. The primary question used for adaptive allocation was: [The patient] would be a good candidate for a prescription for a selective serotonin reuptake inhibitor (SSRI)/referral for a formal ADHD assessment (1 = Strongly Disagree; 2 = Disagree; 3 = Neither disagree nor agree; 4 = Agree; 5 = Strongly agree).

Table 3 | Participant Characteristics (*N* = 420)

	Depression vignette (*N* = 210)	ADHD vignette (*N* = 210)	*p* value
Age, years			
Mean (SD)	49.5 (12.6)	49.6 (12.4)	0.937
Median [Min, Max]	49.0 [28.0, 75.0]	49.0 [28.0, 75.0]	
Missing	3 (0.7%)	3 (0.7%)	
Sex			
Female	78 (37.1%)	100 (47.6%)	0.038
Male	131 (62.4%)	110 (52.4%)	
Decline to state	1 (0.5%)	0 (0%)	
Gender			
Woman	80 (38.1%)	100 (47.6%)	NA
Man	128 (61.0%)	110 (52.4%)	
Nonbinary or Genderqueer	1 (0.5%)	0 (0%)	
Decline to state	1 (0.5%)	0 (0%)	
Ethnicity			
Not Hispanic or Latino	197 (93.8%)	193 (91.9%)	0.651
Hispanic or Latino	9 (4.3%)	13 (6.2%)	
Decline to state	4 (1.9%)	4 (1.9%)	
Race			
American Indian/Alaska Native	2 (1.0%)	2 (1.0%)	0.286
Asian	44 (21.0%)	59 (28.1%)	
Black or African American	8 (3.8%)	9 (4.3%)	
Native Hawaiian or Other Pacific Islander	1 (0.5%)	0 (0%)	
White	130 (61.9%)	126 (60.0%)	
Multiracial	11 (5.2%)	5 (2.4%)	
Other	7 (3.3%)	2 (1.0%)	
Decline to state	7 (3.3%)	7 (3.3%)	
Degree			
M.D.	180 (85.7%)	187 (89.0%)	0.538
M.D./ Ph.D.	3 (1.4%)	4 (1.9%)	
D.O.	26 (12.4%)	19 (9.0%)	
D.O./ Ph.D.	1 (0.5%)	0 (0%)	
Training			
Medical school - USA	168 (80.0%)	167 (79.5%)	NA
Medical school - Abroad	39 (18.6%)	41 (19.5%)	
Post-graduate - USA	29 (13.8%)	30 (14.3%)	
Post-graduate - Abroad	3 (1.4%)	7 (3.3%)	
Other	0 (0%)	1 (0.5%)	
Decline to state	1 (0.5%)	0 (0%)	
Practice years			
Mean (SD)	18.4 (12.3)	18.2 (11.8)	0.848
Median [Q1, Q3]	18.0 [8.0, 26.0]	17.00 [8.0, 26.0]	
Missing	1 (0.5%)	0 (0%)	
Primary subspecialty			
Family medicine	108 (51.4%)	108 (51.4%)	1
Internal medicine	102 (48.6%)	102 (48.6%)	
Practice Region			
Northeast	57 (27.1%)	37 (17.6%)	0.281
Southeast	43 (20.5%)	51 (24.3%)	
Midwest	47 (22.4%)	50 (23.8%)	
Southwest	14 (6.7%)	19 (9.0%)	
West	45 (21.4%)	50 (23.8%)	
Decline to state	4 (1.9%)	3 (1.4%)	
Formal training/education in computer science			
Yes	49 (23.3%)	32 (15.2%)	0.048
No	161 (76.7%)	178 (84.8%)	
Type of computer science training/education			
High school course(s)	22 (10.5%)	12 (5.7%)	NA
Undergraduate course(s)	30 (14.3%)	23 (11.0%)	
Graduate-level course(s)	7 (3.3%)	5 (2.4%)	
CS/Coding bootcamp	5 (2.4%)	1 (0.5%)	
Professional development course(s) or workshop(s)	12 (5.7%)	7 (3.3%)	
Other	2 (1.0%)	3 (1.4%)	
Familiarity with AI/ML			
Not at all familiar	14 (6.7%)	11 (5.2%)	0.486
A little familiar	75 (35.7%)	92 (43.8%)	
Somewhat familiar	96 (45.7%)	83 (39.5%)	
Quite familiar	24 (11.4%)	22 (10.5%)	
Extremely familiar	1 (0.5%)	2 (1.0%)	
Use of AI/ML in research			
Yes	20 (9.5%)	17 (8.1%)	0.773
No	116 (55.2%)	113 (53.8%)	
I do not engage in medical research	74 (35.2%)	80 (38.1%)	
Use of AI/ML in clinical practice			
Yes	49 (23.3%)	31 (14.8%)	0.035
No	161 (76.7%)	179 (85.2%)	
Purpose of AI/ML use in clinical practice			
Shared decision-making	13 (6.2%)	4 (1.9%)	NA
Patient monitoring	11 (5.2%)	13 (6.2%)	
Outcome prediction	15 (7.1%)	10 (4.8%)	
Prediction of complications	4 (1.9%)	5 (2.4%)	
Interpretation/quantification of imaging	8 (3.8%)	5 (2.4%)	
Grading of disease severity	5 (2.4%)	4 (1.9%)	
Diagnosing	25 (11.9%)	15 (7.1%)	
Other	15 (7.1%)	4 (1.9%)	
Current role			
Owner or employee of practice	166 (79.0%)	153 (72.9%)	NA
Employee of an academic institution	17 (8.1%)	28 (13.3%)	
Both Owner or employee of practice and Employee of an academic institution	13 (6.2%)	15 (7.1%)	
Other	11 (5.2%)	12 (5.7%)	
Decline to state	4 (1.9%)	3 (1.4%)	

Participants were allowed to select more than 1 answer for Gender, Training, Type of computer science training/education, Purpose of AI/ML use in clinical practice, and Current role, and therefore, their total percentages may be greater than 100%.

p values are from *t*-tests for continuous variables and Chi-squared tests or Fisher's exact tests as appropriate for categorical variables.

Time point 2. At the second time point, physicians were presented with a subvignette that described a hypothetical AI CDSS that was intended to assist the physician in their assessment of the patient. Physicians were randomized to receive either a brief or a longer explanation regarding the AI system. They were then presented with 9 questions about their perception and attitudes towards the AI system based on the provided amount of information. The primary question used for adaptive allocation was: If given the option, I

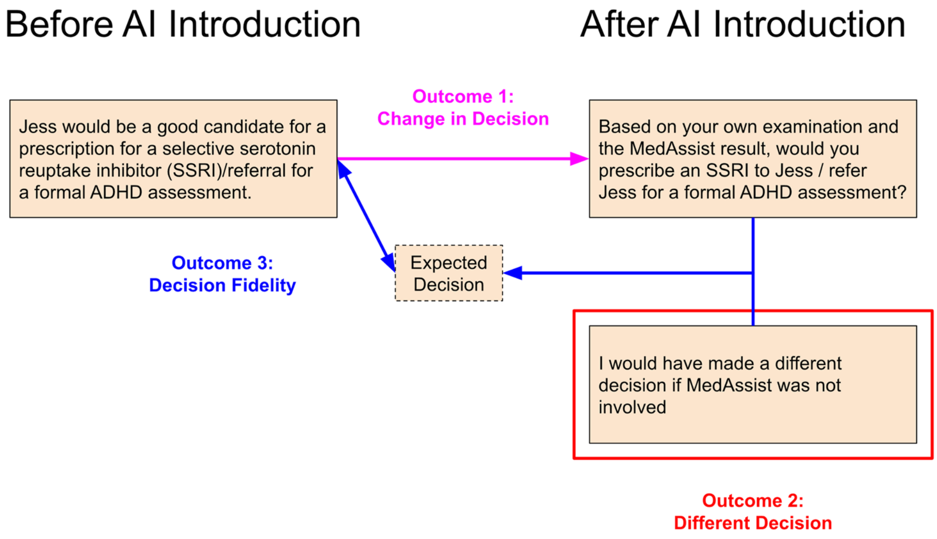
Before AI Introduction
After AI Introduction
Jess would be a good candidate for a prescription for a selective serotonin reuptake inhibitor (SSRI)/referral for a formal ADHD assessment.
Outcome 1: Change in Decision
Based on your own examination and the MedAssist result, would you prescribe an SSRI to Jess / refer Jess for a formal ADHD assessment?
Outcome 3: Decision Fidelity
Expected Decision
I would have made a different decision if MedAssist was not involved
Outcome 2: Different Decision

Fig. 3 | Diagram of main and exploratory outcomes. The figure depicts the question items before and after introducing AI to the scenarios. For exploratory outcome 1, the change in decision was obtained by the difference between the physicians' initial assessment and their final decision (denoted in magenta). The outcome "I would have made a different decision if MedAssist was not involved" (outlined in red) was the second outcome. Decision fidelity outcome was obtained by comparing the initial decision and expected decision. The Dotted box indicates the intermediate variable calculated from the final observed decision.

would choose to use [the AI tool] in my practice (1 = Strongly Disagree; 2 = Disagree; 3 = Neither disagree nor agree; 4 = Agree; 5 = Strongly agree).

Time point 3. At the third time point, physicians were presented with a subvignette describing the output of the AI system. Physicians were randomized to receive either higher-certainty output which recommended further treating the patient, or a lower-certainty output which recommended not treating the patient further at this time. They were then asked seven questions regarding their final clinical decision, related confidence, and the perceived impact of AI on their final decision. The designated primary question was: Based on your own examination and the [AI] result, would you prescribe an SSRI to [the patient]/refer [the patient] for a formal ADHD assessment? (Yes; No).

Statistical analysis

Question items. Three question items from the vignette survey were used in our analysis, one item before the introduction of AI in the scenarios (time 1), and two items after its introduction (time 3).

Outcomes. The first outcome "change in decision" is a binary variable obtained by taking the absolute value of the difference between the physician assessment at time 1 and 3 (see Fig. 3). The second outcome was the observed response to the item "I would have made a different decision if AI was not involved."

Exploratory outcome. We defined a new exploratory outcome as "fidelity" of the physician's decision. This was a binary outcome for whether there was consistency between the initial assessment without AI and the "expected decision." The "expected decision" was an intermediate variable that we defined based on the self-reported response to the item "I would have made a different decision": if the response to this item was "yes," then we took the observed final decision and switched the response, if the answer to this item was "no", then we kept the observed decision as it was reported.

Predictors. Vignette characteristics i.e., clinical decision type, amount of information on the AI, and certainty of the AI output, were used as predictors. Familiarity with AI was also included. This binary independent variable was defined to be om if they reported researching AI, using AI in clinical practice, or having familiarity with AI in the baseline survey, and 0 otherwise.

Analyses. Regression models involving covariates and interaction effects were fitted for each outcome. Logistic regression models were fitted, and the corresponding conditional odds ratios and the associated confidence intervals were used to conduct the exploratory analyses. For the first exploratory analysis, the outcome for change in decision was regressed on dichotomous initial clinical assessment, AI recommendation, and their interaction. For the second exploratory analysis, we focused on physicians' reports on whether they would have made a different decision without AI; this outcome was dichotomized and regressed on their initial decision, the output of AI, as well as their interaction. Lastly, for the third exploratory analysis, the fidelity variable was regressed on the same variables as the second analysis but with familiarity with AI included as an additional predictor. Odds ratios with 95% confidence intervals that did not include 1 were considered statistically significant.

Data availability

Restrictions apply to the availability of the data that support the findings of this study in order to protect study participant privacy, and thus are not publicly available.

Code availability

The free programming languages R (3.6.3) and Python (3.7.3) were used to perform all statistical analyses.

Received: 22 June 2024; Accepted: 11 March 2025;
Published online: 10 May 2025

References

1. Lin, S. A clinician's guide to artificial intelligence (AI): why and how primary care should lead the health care ai revolution. *J. Am. Board Fam. Med.* **35**, 175–184 (2022).
2. Jain, A. et al. Development and assessment of an artificial intelligence–based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. *JAMA Netw. Open* **4**, e217249 (2021).
3. Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N. & Folk, J. C. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digit. Med.* **1**, 39 (2018).
4. Jones, O. T. et al. Artificial intelligence techniques that may be applied to primary care data to facilitate earlier diagnosis of cancer: systematic review. *J. Med. Internet Res.* **23**, e23483 (2021).
5. Kueper, J. K., Terry, A. L., Zwarenstein, M. & Lizotte, D. J. Artificial intelligence and primary care research: a scoping review. *Ann. Fam. Med.* **18**, 250–258 (2020).
6. Lee, E. E. et al. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **6**, 856–864 (2021).

7. Regier, D. The de facto US mental health services system. *Arch. Gen. Psychiatry* **35**, 685 (1978).
8. Chen, Z. S. et al. Modern views of machine learning for precision psychiatry. *Patterns* **3**, 100602 (2022).
9. Barua, P. D. et al. Artificial intelligence assisted tools for the detection of anxiety and depression leading to suicidal ideation in adolescents: a review. *Cogn. Neurodyn.* **18**, 1–22 (2024).
10. Sau, A. & Bhakta, I. Predicting anxiety and depression in elderly patients using machine learning technology. *Healthc. Technol. Lett.* **4**, 238–243 (2017).
11. Kannampallil, T. et al. Cross-trial prediction of depression remission using problem-solving therapy: a machine learning approach. *J. Affect. Disord.* **308**, 89–97 (2022).
12. Garriga, R. et al. Machine learning model to predict mental health crises from electronic health records. *Nat. Med.* **28**, 1240–1248 (2022).
13. Liu, C.-F., Chen, Z.-C., Kuo, S.-C. & Lin, T.-C. Does AI explainability affect physicians' intention to use AI?. *Int. J. Med. Inf.* **168**, 104884 (2022).
14. Diprose, W. K. et al. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *J. Am. Med. Inform. Assoc. JAMIA* **27**, 592–600 (2020).
15. Gaube, S. et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *npj Digit. Med.* **4**, 1–8 (2021).
16. Kiani, A. et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *npj Digit. Med.* **3**, 1–8 (2020).
17. Agarwal, N., Moehring, A., Rajpurkar, P. & Salz, T. *Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology*. https://doi.org/10.3386/w31422 (2023).
18. 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.* **32**, 713–735 (2021).
19. Liberati, E. G. et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implement. Sci.* **12**, 113 (2017).
20. Liyanage, H. et al. Artificial intelligence in primary health care: perceptions, issues, and challenges. *Yearb. Med. Inform.* **28**, 41–46 (2019).
21. Kim, J. P. et al. Physicians' and machine learning researchers' perspectives on ethical issues in the early development of clinical machine learning tools: qualitative interview study. *JMIR AI* **2**, e47449 (2023).
22. Kim, J. P. & Yang, H.-J. A novel experimental vignette methodology: SMART vignettes. *Methodol. Innov*. https://doi.org/10.1177/20597991241240081 (2024).
23. Kim, J. P., Yang, H.-J., Kim, B., Ryan, K. & Roberts, L. W. Understanding physician's perspectives on ai in health care: protocol for a sequential multiple assignment randomized vignette study. *JMIR Res. Protoc.* **13**, e54787 (2024).
24. Tobia, K., Nielsen, A. & Stremitzer, A. When does physician use of AI increase liability?. *J. Nucl. Med.* **62**, 17–21 (2021).
25. Banja, J. D., Hollstein, R. D. & Bruno, M. A. When artificial intelligence models surpass physician performance: medical malpractice liability in an era of advanced artificial intelligence. *J. Am. Coll. Radiol.* **19**, 816–820 (2022).
26. Bernstein, M. H. et al. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. *Eur. Radiol.* **33**, 8263–8269 (2023).
27. Wysocki, O. et al. Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI-driven clinical decision-making. *Artif. Intell.* **316**, 103839 (2023).
28. Nagendran, M., Festor, P., Komorowski, M., Gordon, A. C. & Faisal, A. A. Quantifying the impact of AI recommendations with explanations on prescription decision making. *npj Digit. Med.* **6**, 1–7 (2023).
29. Clement, J., Ren, Y. & Curley, S. AI quality and clinical roles impact decision quality in AI-Augmented medical decisions more than AI explanations. https://doi.org/10.2139/ssrn.3961156 (2021).
30. Markus, A. F., Kors, J. A. & Rijnbeek, P. R. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J. Biomed. Inform.* **113**, 103655 (2021).
31. American Medical Association. *AMA Physician Professional DataTM*. https://www.ama-assn.org/about/physician-professional-data/ama-physician-professional-data (American Medical Association, 2023).

Acknowledgements

This study was funded by the National Center for Advancing Translational Sciences [R01 TR003505]. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author contributions

K.R.: Methodology, Writing - Original Draft, Writing - Review and Editing, Visualization; H.Y.: Methodology, Formal analysis, Writing - Original Draft, Writing - Review and Editing, Visualization; B.K.: Formal analysis, Writing - Original Draft; J.P.K.: Conceptualization, Methodology, Supervision, Writing - Original Draft, Writing - Review and Editing. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44184-025-00124-y.

Correspondence and requests for materials should be addressed to Jane Paik Kim.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.