

# **Patient reactions to AI-clinician discrepancies: Web-based randomized experiment**

Farrah Madanay, Laura S. O'Donohue, Brian J. Zikmund-Fisher

Submitted to: Journal of Medical Internet Research  
on: November 14, 2024

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## Table of Contents

---

|                                  |    |
|----------------------------------|----|
| <b>Original Manuscript</b> ..... | 5  |
| <b>Supplementary Files</b> ..... | 22 |
| Figures.....                     | 23 |
| Figure 1.....                    | 24 |
| Figure 3.....                    | 26 |
| Multimedia Appendixes.....       | 28 |
| Multimedia Appendix 1.....       | 29 |

# Patient reactions to AI-clinician discrepancies: Web-based randomized experiment

Farrah Madanay<sup>1</sup> PhD; Laura S. O'Donohue<sup>2</sup> MD; Brian J. Zikmund-Fisher<sup>3</sup> PhD

<sup>1</sup> Center for Bioethics and Social Sciences in Medicine University of Michigan–Ann Arbor Ann Arbor US

<sup>2</sup> Department of Radiology University of Michigan Medicine University of Michigan–Ann Arbor Ann Arbor US

<sup>3</sup> Health Behavior and Health Equity, Internal Medicine, Center for Bioethics and Social Sciences in Medicine University of Michigan–Ann Arbor Ann Arbor US

## Corresponding Author:

Farrah Madanay PhD

Center for Bioethics and Social Sciences in Medicine  
University of Michigan–Ann Arbor  
2800 Plymouth Rd., Bldg. 14, G016  
Ann Arbor  
US

## Abstract

**Background:** As FDA-approved AI for medical imaging rises, radiologists are increasingly integrating AI into their clinical practices. In lung cancer screening, diagnostic AI offers a second set of eyes with the potential to detect cancer earlier than human radiologists. Despite AI's promise, a potential problem with its integration is the erosion of patient confidence in clinician expertise when there is a discrepancy between the radiologist's and the AI's interpretation of the imaging test.

**Objective:** This study aims to examine how discrepancies between AI-derived recommendations and radiologists' recommendations affect patients' agreement with radiologists' recommendations and satisfaction in their radiologists. We also analyze how patients' medical maximizing-minimizing (MMM) preferences moderate these relationships.

**Methods:** We conducted a randomized, between-subjects experiment with 1,606 U.S. adult participants. Assuming the role of patients, participants imagined undergoing a low-dose computerized tomography scan for lung cancer screening and receiving results and recommendations from: (1) a radiologist only, (2) AI and a radiologist in agreement, (3) a radiologist who recommended more testing than AI (i.e., radiologist overcalled AI), or (4) a radiologist who recommended less testing than AI (i.e., radiologist undercalled AI). Participants rated the radiologist on three criteria: agreement with the radiologist's recommendation, how likely they would be to recommend the radiologist to family and friends, and how good of a provider they perceived the radiologist to be. We also measured MMM and categorized participants as maximizers (i.e., those who seek aggressive intervention), minimizers (i.e., those who prefer no or passive intervention), and neutrals (i.e., those in the middle).

**Results:** Participants' agreement with the radiologist's recommendation was significantly lower when the radiologist undercalled AI ( $M=4.01$ ,  $SE=0.07$ ) vs. overcalled ( $M=4.63$ ,  $SE=0.06$ ), agreed with ( $M=4.55$ ,  $SE=0.07$ ), or had no AI ( $M=4.57$ ,  $SE=0.06$ ;  $P<.001$ ). Additionally, participants were least likely to recommend ( $P<.001$ ) and positively rate ( $P<.001$ ) the radiologist who undercalled AI. There were no significant differences among the other conditions. Maximizers strongly agreed ( $b=0.82$ ,  $SE=0.14$ ;  $P<.001$ ) with the radiologist who overcalled AI whereas minimizers strongly disagreed ( $b=-0.43$ ,  $SE=0.18$ ,  $P=.015$ ), compared to neutrals. Moreover, maximizers strongly disagreed with the radiologist who undercalled AI ( $b=-0.47$ ,  $SE=0.14$ ;  $P=.001$ ).

**Conclusions:** Radiologists who recommend less testing than AI may face decreased patient confidence in their expertise, but they may not face this same penalty for giving more aggressive recommendations than AI. Patients' reactions may depend in part on whether their general preferences to maximize or minimize align with the radiologists' recommendations.

(JMIR Preprints 14/11/2024:68823)

DOI: <https://doi.org/10.2196/preprints.68823>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone.

## Original Manuscript

## Original Paper

## Patient reactions to AI-clinician discrepancies: Web-based randomized experiment

**ABSTRACT**

*Background:* As FDA-approved AI for medical imaging rises, radiologists are increasingly integrating AI into their clinical practices. In lung cancer screening, diagnostic AI offers a second set of eyes with the potential to detect cancer earlier than human radiologists. Despite AI's promise, a potential problem with its integration is the erosion of patient confidence in clinician expertise when there is a discrepancy between the radiologist's and the AI's interpretation of the imaging test.

*Objective:* This study aims to examine how discrepancies between AI-derived recommendations and radiologists' recommendations affect patients' agreement with radiologists' recommendations and satisfaction in their radiologists. We also analyze how patients' medical maximizing-minimizing (MMM) preferences moderate these relationships.

*Methods:* We conducted a randomized, between-subjects experiment with 1,606 U.S. adult participants. Assuming the role of patients, participants imagined undergoing a low-dose computerized tomography scan for lung cancer screening and receiving results and recommendations from: (1) a radiologist only, (2) AI and a radiologist in agreement, (3) a radiologist who recommended more testing than AI (i.e., radiologist overcalled AI), or (4) a radiologist who recommended less testing than AI (i.e., radiologist undercalled AI). Participants rated the radiologist on three criteria: agreement with the radiologist's recommendation, how likely they would be to recommend the radiologist to family and friends, and how good of a provider they perceived the radiologist to be. We measured MMM and categorized participants as maximizers (i.e., those who seek aggressive intervention), minimizers (i.e., those who prefer no or passive intervention), and neutrals (i.e., those in the middle).

*Results:* Participants' agreement with the radiologist's recommendation was significantly lower when the radiologist undercalled AI ( $M=4.01$ ,  $SE=0.07$ ,  $P<.001$ ) compared to the other three conditions, between which there were no significant differences (radiologist overcalled AI [ $M=4.63$ ,  $SE=0.06$ ], agreed with AI [ $M=4.55$ ,  $SE=0.07$ ], or had no AI [ $M=4.57$ ,  $SE=0.06$ ]). Similarly, participants were least likely to recommend ( $P<.001$ ) and positively rate ( $P<.001$ ) the radiologist who undercalled AI, with no significant differences among the other conditions. Maximizers agreed with the radiologist who overcalled AI ( $b=0.82$ ,  $SE=0.14$ ;  $P<.001$ ) and disagreed with the radiologist who undercalled AI ( $b=-0.47$ ,  $SE=0.14$ ;  $P=.001$ ). However, whereas minimizers disagreed with the radiologist who overcalled AI ( $b=-0.43$ ,  $SE=0.18$ ,  $P=.02$ ), they did not significantly agree with the radiologist who undercalled AI ( $b=0.14$ ,  $SE=0.17$ ,  $P=.41$ ).

*Conclusions:* Radiologists who recommend less testing than AI may face decreased patient confidence in their expertise, but they may not face this same penalty for giving more aggressive recommendations than AI. Patients' reactions may depend in part on whether their general preferences to maximize or minimize align with the radiologists' recommendations. Future research should test communication strategies for radiologists' disclosure of AI discrepancies to patients.

*Keywords:* artificial intelligence; medical maximizing-minimizing; early detection of cancer; decision making; radiologists; patient satisfaction; patient-physician relationship; communication

**INTRODUCTION**

Radiologists are trained to detect lung cancer from low-dose computerized tomography (LDCT) scans with high accuracy [1, 2] and medical artificial intelligence (AI) promises to improve this accuracy detection by acting as a “second pair of eyes” [1, 3, 4]. As of August 2024, the U.S. Food and Drug Administration (FDA) published a list of 950 authorized AI-enabled medical devices, 723 (76%) of which were in the field of radiology. The developers of one medical AI model meant for lung cancer screening, called Sybil, claim the AI can detect future lung cancer in CT scans up to 6 years before lesions become visible to human radiologists [5]. In a 2020 survey of American College of Radiology members, 33% of radiologists reported currently using AI, while 20% of those not currently using AI planned to use AI in the next 1 to 5 years [6].

Despite this rise in medical AI among radiologists, patient acceptance lags. Prior research has found patients’ resistance to AI stems from beliefs that AI performs inferior to humans and cannot be held accountable for errors, lower objective and subjective understanding of AI versus human decision-making, and fear that AI does not meet patients’ unique needs [7-10].

In part due to concerns about AI accuracy, U.S. federal policies emphasize including a “human in the loop” (HITL) when it comes to AI development and use. In 2021, the FDA, along with Health Canada and the United Kingdom’s Medicines and Healthcare Products Regulatory Agency, published 10 guiding principles for machine learning development [11]. One of the principles describes focusing on the performance of the “human-AI team,” with AI performing with a HITL rather than in isolation.

Despite these guidelines, little is known about how patients will respond to the HITL. Whereas prior research has focused on patients’ trust or mistrust of medical AI [7, 8, 10], few studies have examined patients’ trust in the clinician of the human-AI team. Research shows patients’ trust in their clinician is related to patient satisfaction, compliance, and health outcomes [12-14]. Yet, medical AI could fundamentally change the patient-clinician relationship, and erode patients’ trust in clinician expertise [15, 16].

In this study, we ask, how do human-AI discrepancies affect patients’ agreement with radiologist recommendations and satisfaction in their radiologist? We examine two directions of human-AI discrepancy. First, the radiologist can “overcall AI,” in which the radiologist identifies higher cancer risk than the AI and recommends more testing. Second, the radiologist can “undercall AI,” in which the radiologist identifies lower cancer risk than the AI and recommends less testing. We hypothesize that patients will least agree with the recommendation of the radiologist who undercalls AI.

We anticipate patients’ agreement with their radiologist’s recommendation will vary based on individual differences, like medical maximizing-minimizing (MMM) preferences. Whereas medical maximizers typically seek more aggressive and optional approaches to care, minimizers prefer no or passive medical intervention unless deemed completely necessary [17]. Researchers have shown that MMM is associated with preferences for cancer screening and surveillance [18, 19], treatment of incidental findings on imaging tests [20], concern with stopping medication [21], pursuit of appropriate care [22], and avoidance of healthcare [23]. In this context, we hypothesize that measuring MMM might help to identify opportunities to tailor communication and guide shared decision-making about medical AI to patients with different underlying care preferences.

## METHODS

### Participants

### and

### Procedure

This study was preregistered at Open Science Framework: <https://osf.io/j3yf2>. We programmed the online quantitative experimental questionnaire in Qualtrics (Qualtrics, Provo, UT) and collected data in April and May 2024. A total of 1828 adult, English-speaking participants were recruited by Dynata (<https://www.dynata.com>). Dynata facilitates data collection for research surveys through its panel of potential online participants. We provided Dynata with quotas for age, gender, and race/ethnicity based on the U.S. population composition. Inclusion criteria were at least 18 years of age and

residents of the U.S. These quotas and inclusion criteria enabled us to receive a diverse range of participant reactions to our experiment.

## Study

The study was designed as a one-way, between-subjects, randomized experiment. Participants assumed the role of a patient advised to undergo a low-dose CT (LDCT) scan for lung cancer screening. After reading the hypothetical scenario, participants were randomized to receive one of four results conditions, in which (1) a radiologist alone identified low risk of cancer and recommended a repeat screening CT in 6 months, (2) AI and a radiologist both identified low risk of cancer and recommended a repeat screening CT in 6 months, (3) AI identified medium risk of cancer and recommended immediate additional testing with a nuclear medicine exam (positron emission tomography/computed tomography [PET/CT]) but a radiologist identified low risk and recommended a repeat screening CT in 6 months (i.e., radiologist undercalled AI), or (4) AI identified low risk of cancer and recommended a repeat screening in 6 months but a radiologist identified medium risk and recommended immediate additional testing (PET/CT; i.e., radiologist overcalled AI). We used the randomizer within Qualtrics to randomly assign participants to conditions. Participants read the results both in an electronic health record (EHR) system report and in a transcript of an oral follow-up. The study was declared exempt from review by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board. *Figure 1* shows the LDCT scan results received in the four experimental conditions.

## Design

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p style="text-align: center;"><b>Radiologist only</b></p> <p><u>IMPRESSION:</u></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p style="padding-left: 40px;">Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><u>RECOMMENDATION:</u></p> <p>Repeat low-dose CT in 6 months to ensure no interval growth.</p>                                                                                                                                                                                                                                                                                                                                                      | <p style="text-align: center;"><b>Radiologist-AI agreement</b></p> <p><u>IMPRESSION:</u></p> <p><b>LungDetect AI:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p style="padding-left: 40px;">Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><b>Radiologist:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p style="padding-left: 40px;">Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><u>RECOMMENDATION:</u></p> <p>Repeat low-dose CT in 6 months to ensure no interval growth, based on AI and radiologist agreement.</p>                                                                  |
| <p style="text-align: center;"><b>Radiologist overcalls AI</b></p> <p><u>IMPRESSION:</u></p> <p><b>LungDetect AI:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p style="padding-left: 40px;">Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><b>Radiologist:</b></p> <p>1. There is a 9-millimeter solid nodule in the left lung.</p> <p style="padding-left: 40px;">Lung-RADS 4A: Suspicious. <b>Medium risk of malignancy (cancer) 5-15%</b></p> <p><u>RECOMMENDATION:</u></p> <p><b>LungDetect AI:</b></p> <p>Repeat low-dose CT in 6 months to ensure no interval growth.</p> <p><b>Radiologist:</b></p> <p>Undergo additional imaging immediately.</p> | <p style="text-align: center;"><b>Radiologist undercalls AI</b></p> <p><u>IMPRESSION:</u></p> <p><b>LungDetect AI:</b></p> <p>1. There is a 9-millimeter solid nodule in the left lung.</p> <p style="padding-left: 40px;">Lung-RADS 4A: Suspicious. <b>Medium risk of malignancy (cancer) 5-15%</b></p> <p><b>Radiologist:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p style="padding-left: 40px;">Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><u>RECOMMENDATION:</u></p> <p><b>LungDetect AI:</b></p> <p>Undergo additional imaging immediately.</p> <p><b>Radiologist:</b></p> <p>Repeat low-dose CT in 6 months to ensure no interval growth.</p> |

Figure 1: Hypothetical results of the LDCT scan, formatted to align with common EHR systems,

received by participants, by condition.

Participants chose their preferred next exam (repeat CT in 6 months or PET/CT now) and rated how much they agreed with the radiologist's recommendation. They also rated their satisfaction with the radiologist. Participants then answered questions about their attitudes toward medical AI and their MMM preferences. See *Supplementary Materials* for the hypothetical scenario and questionnaire measures.

The main dependent variable was a single, 6-point Likert scale measure asking participants how much they agreed or disagreed with the radiologist's recommendation (1 = strongly disagree, 6 = strongly agree). We also measured participants' satisfaction with their radiologist using questions adapted from the Agency for Healthcare Research and Quality's Consumer Assessment of Healthcare Providers and Systems patient-experience surveys [24]. These included two 10-point scales asking participants how likely they would be to recommend the radiologist to family and friends and how good of a provider they perceived the radiologist.

Covariates included mean composite scores of participants' attitudes toward medical AI and of participants' MMM preferences. Attitudes toward medical AI were measured by asking participants to rate their agreement with 7 statements on 6-point Likert scales (1 = strongly disagree, 6 = strongly agree). Five of these statements were adapted from prior literature [25, 26], and two were added by the authors. We reverse-coded statements 2 through 6 before creating the composite so that higher scores represented greater affinity for medical AI. Cronbach's alpha for the seven measures was  $\alpha = 0.62$ , indicating acceptable internal consistency.

We measured MMM using two previously published questions [17, 27]. The measures asked, in medical situations, whether participants: (1) tended to lean toward waiting and seeing or taking action (1 = I strongly lean toward waiting and seeing, 6 = I strongly lean toward taking action) [17], and (2) tended to lean toward doing only what is necessary or everything possible (1 = I strongly lean toward doing only what is necessary, 6 = I strongly lean toward doing everything possible) [27]. Higher scores on these measures indicated greater preference for medical maximizing. Cronbach's alpha for the two measures was  $\alpha = 0.74$ , indicating good internal consistency. We also categorized participants as minimizers (1–2.5), neutral (3–4), and maximizers (4.5–6), in line with prior research [27, 28], for ease of interpretation.

## Analysis

We conducted ordinary least squares (OLS) regressions to test whether receiving a recommendation from a radiologist alone, or in agreement or disagreement with AI affected participants' agreement with their radiologist's recommendation. Two-sided  $P < .05$  was considered significant. In exploratory analyses, we also tested how experimental condition affected participants' follow-up test choice and participants' satisfaction with their radiologist.

In sensitivity analyses, we conducted Kruskal-Wallis H tests and Dunn's tests and ordinal logistic regression models to check the robustness of the effects on our main dependent variable, agreement with the radiologist's recommendation. We also conducted regressions with the inclusion of medical AI attitudes, MMM preferences, and participant demographics as covariates to check the sensitivity of the findings to the inclusion of controls. We ran moderation analyses with participants' MMM preferences first as a continuous variable and second as a categorical variable.

## RESULTS

### Descriptive data

The average time to completion was 8 minutes and 32 seconds after excluding 222 speeder participants who completed the questionnaire in 3 minutes or less. A total of 1606 (88%) participants

(51% female, mean age=51.49 y) were used in the analyses. Participants in each condition were similar in mean age and sex, indicating successful randomization (*Table 1*).



Table 1: Descriptive statistics for the sample (N=1606)

| Characteristic                                      | No. (%)                        |                                            |                                        |                                         | Total<br>(N=1606) | P<br>value <sup>a</sup> |
|-----------------------------------------------------|--------------------------------|--------------------------------------------|----------------------------------------|-----------------------------------------|-------------------|-------------------------|
|                                                     | Radiologist<br>only<br>(n=413) | Radiologist<br>-AI<br>agreement<br>(n=401) | Radiologist<br>overcalls<br>AI (n=408) | Radiologist<br>undercalls<br>AI (n=384) |                   |                         |
| <b>Age, mean</b>                                    | 52.15                          | 50.79                                      | 51.29                                  | 51.72                                   | 51.49             | .69                     |
| <b>Gender</b>                                       |                                |                                            |                                        |                                         |                   |                         |
| Female                                              | 215 (52)                       | 206 (51)                                   | 219 (54)                               | 183 (48)                                | 823 (51)          | .67                     |
| Male                                                | 194 (47)                       | 192 (48)                                   | 187 (46)                               | 199 (52)                                | 772 (48)          |                         |
| Other                                               | 4 (1)                          | 3 (1)                                      | 2 (0)                                  | 2 (1)                                   | 11 (1)            |                         |
| <b>Highest level of<br/>completed<br/>education</b> |                                |                                            |                                        |                                         |                   |                         |
| HS or less                                          | 80 (19)                        | 97 (24)                                    | 90 (22)                                | 77 (20)                                 | 344 (21)          | .28                     |
| Some<br>postsecondary                               | 133 (32)                       | 127 (32)                                   | 117 (29)                               | 137 (36)                                | 514 (32)          |                         |
| Bachelor's                                          | 125 (30)                       | 105 (26)                                   | 127 (31)                               | 108 (28)                                | 465 (29)          |                         |
| Graduate                                            | 73 (18)                        | 67 (17)                                    | 74 (18)                                | 60 (16)                                 | 274 (17)          |                         |
| Other                                               | 2 (0)                          | 5 (1)                                      | 0 (0)                                  | 2 (1)                                   | 9 (1)             |                         |
| <b>Race</b>                                         |                                |                                            |                                        |                                         |                   |                         |
| White                                               | 314 (76)                       | 286 (71)                                   | 301 (74)                               | 290 (76)                                | 1191 (74)         | .76                     |
| Black                                               | 41 (10)                        | 55 (14)                                    | 50 (12)                                | 50 (13)                                 | 196 (12)          |                         |
| Asian                                               | 22 (5)                         | 22 (5)                                     | 24 (6)                                 | 17 (4)                                  | 85 (5)            |                         |
| Multi/Other                                         | 36 (9)                         | 38 (9)                                     | 33 (8)                                 | 27 (7)                                  | 134 (8)           |                         |
| <b>Hispanic/<br/>Latino</b>                         |                                |                                            |                                        |                                         |                   |                         |
| Not Hispanic                                        | 332 (80)                       | 314 (78)                                   | 344 (84)                               | 317 (83)                                | 1307<br>(81)      | .42                     |
| Hispanic                                            | 78 (19)                        | 82 (20)                                    | 60 (15)                                | 63 (16)                                 | 283 (18)          |                         |
| Other                                               | 3 (1)                          | 5 (1)                                      | 4 (1)                                  | 4 (1)                                   | 16 (1)            |                         |

**Notes:**

Abbreviations: HS, high school

<sup>a</sup>ANOVA for age, chi-square tests for gender, education, race, and Hispanic/Latino.

Some percentages may total &gt;100 due to rounding.

## Main Results

### Participants' agreement with the radiologist's recommendation

Participants were less likely to agree with the radiologist who undercalled AI ( $M=4.01$ ,  $SE=0.07$ ), which was significantly different than the other conditions ( $P<.001$ ). Participants' agreement with their radiologist's recommendation was not significantly different among the radiologist-only ( $M=4.57$ ,  $SE=0.06$ ), radiologist-AI agreement ( $M=4.55$ ,  $SE=0.07$ ), and radiologist-overcalls-AI ( $M=4.63$ ,  $SE=0.06$ ) conditions (Figure 2). Results were robust to the inclusion of covariates, including medical AI attitudes, MMM preferences, and participant demographics (Supplementary Table 1), and remained consistent when using alternative model specifications, including the Kruskal-Wallis H tests and Dunn's tests and ordinal logistic regressions (Supplementary Tables 2 and 3).

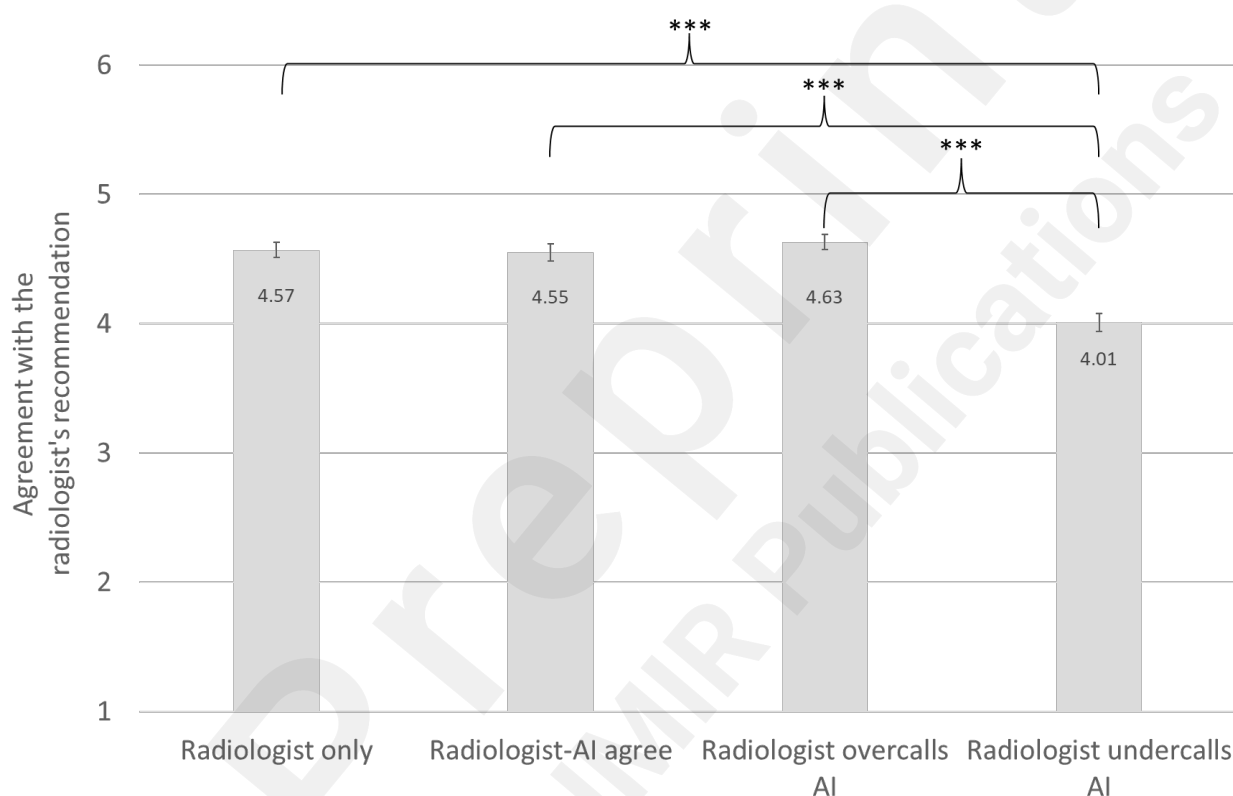


Figure 2: Participants' agreement with the radiologist's recommendation, by condition. Means and standard error bars depicted. Stars represent significant differences between groups ( $***P<.001$ ).

### Moderation by MMM preferences

MMM scores had a strong positive effect on participants' agreement in the radiologist-overcalls-AI condition ( $b=0.42$ ,  $SE=0.05$ ,  $P<.001$ ) and a negative effect on the radiologist-undercalls-AI condition ( $b=-0.21$ ,  $SE=0.05$ ,  $P<.001$ ). There was no significant effect on the radiologist-only ( $b=-0.07$ ,  $SE=0.05$ ,  $P=.13$ ) and radiologist-AI agreement ( $b=-0.04$ ,  $SE=0.05$ ,  $P=.42$ ) conditions (Figure 3).

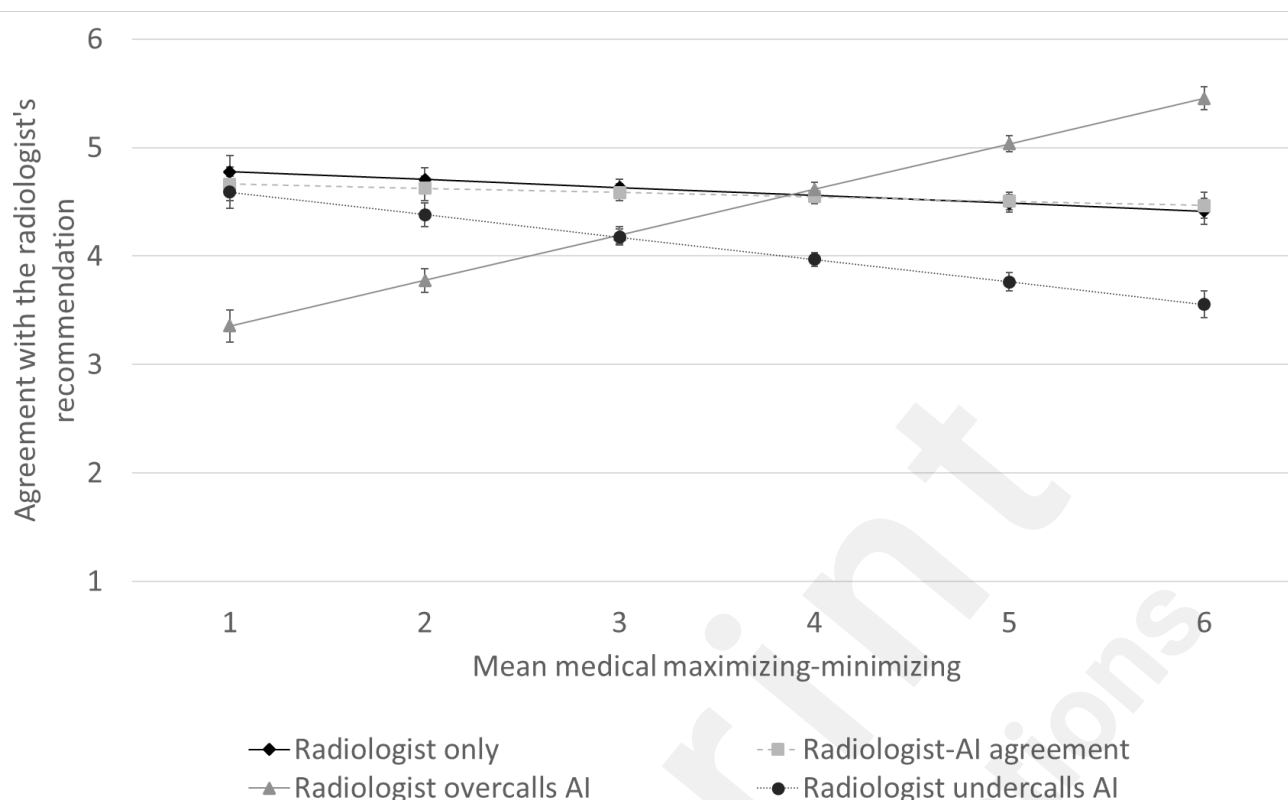


Figure 3: Effect of participants' medical maximizing-minimizing preferences on agreement with the radiologist's recommendation, by condition. Standard error bars depicted.

In our sample, MMM scores were left-skewed; 302 (18.80%) participants were minimizers, 654 (40.72%) were neutrals, and 650 (40.47%) were maximizers. When analyzing by MMM category, we found, relative to neutrals, minimizers disagreed more strongly with the radiologist who overcalled AI ( $b=-0.43$ ,  $SE=0.18$ ,  $P=.02$ ), but did not agree more strongly with radiologists who undercalled AI ( $b=0.14$ ,  $SE=0.17$ ,  $P=.41$ ). Relative to neutrals, maximizers agreed more strongly ( $b=0.82$ ,  $SE=0.14$ ,  $P<.001$ ) with radiologists who overcalled AI and disagreed more strongly when radiologists undercalled AI, ( $b=-0.47$ ,  $SE=0.14$ ,  $P=.001$ ; *Figure 4*). The pattern of this moderation analysis with MMM as a categorical variable matched that of the analysis using MMM as a continuous variable.

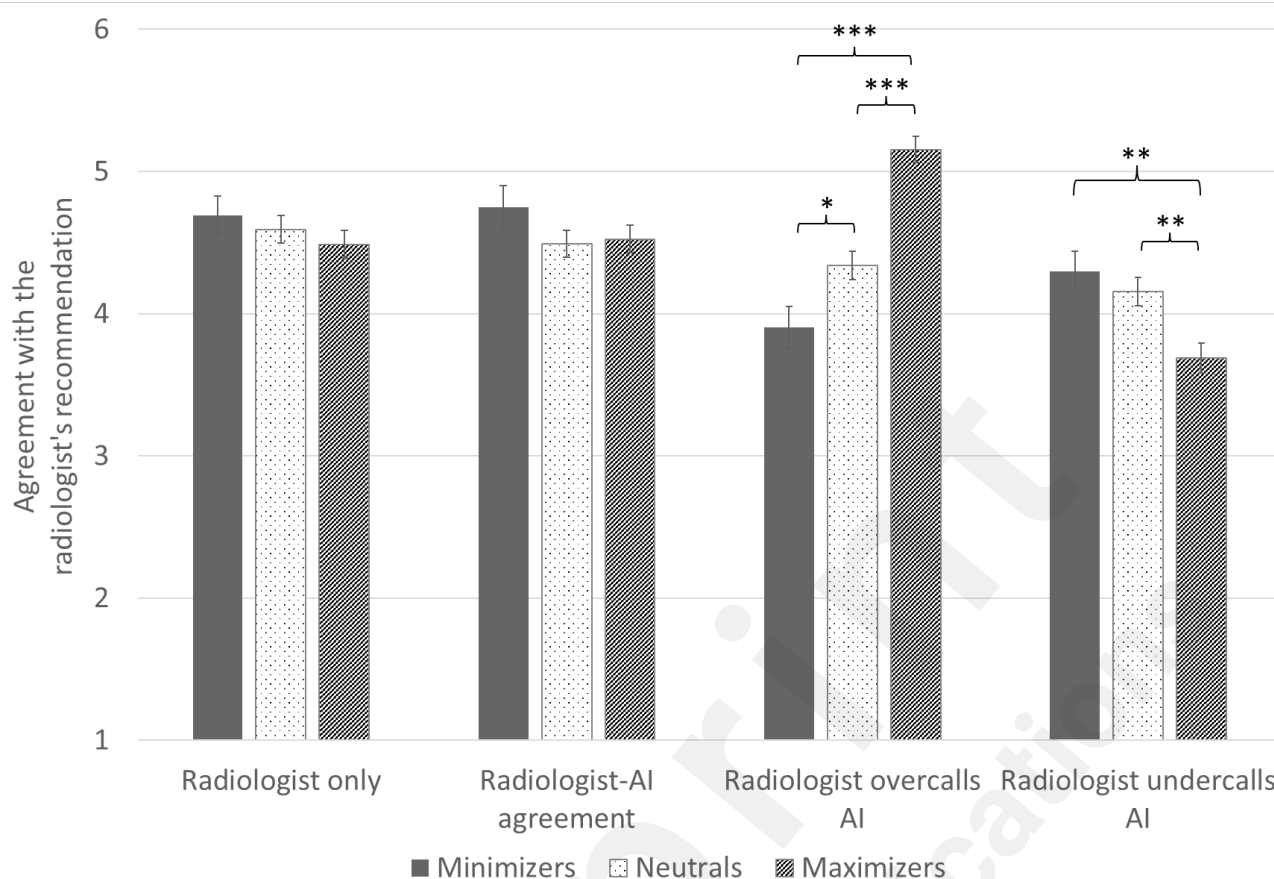


Figure 4: Effect of condition on participants' agreement with the radiologist's recommendation, by participants' MMM category. Means and standard error bars depicted. Stars represent significant differences between groups (\* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ ).

## Secondary Results

### Patients' choice between follow-up screening CT in 6 months or additional testing (PET/CT) immediately

Participants were least likely to follow the recommendation of the radiologist who undercalled AI (48%,  $SE=0.03$ ), which was significantly different than the other conditions ( $P < .001$ ). Participants' choice to follow their radiologist's recommendation was not significantly different among the radiologist-only (63%,  $SE=0.02$ ), radiologist-AI agreement (63%,  $SE=0.02$ ), and radiologist-overcalls-AI conditions (64%,  $SE=0.02$ ; *Supplementary Figure 1*). This pattern aligned with the results of participants' agreement with their radiologist's recommendation.

### Patients' likelihood to recommend the radiologist and radiologist rating

In exploratory analyses, we examined participants' likelihood to recommend their radiologist and rating of their radiologist after receiving the results of their LDCT scan. *Table 2* presents means and standard errors for all dependent variables. Participants were both least likely to recommend ( $M=6.47$ ,  $SE=0.13$ ) and positively rate ( $M=6.82$ ,  $SE=0.11$ ) the radiologist who undercalled AI, which was significantly different from the other conditions ( $P < .001$ ). There were no significant differences among the radiologist-only, radiologist-AI agreement, and radiologist-overcalls-AI conditions in participants' recommendation likelihood ( $M=7.24$ ,  $SE=0.12$  vs.  $M=7.40$ ,  $SE=0.12$  vs.  $M=7.43$ ,  $SE=0.11$ ) or rating of the radiologist ( $M=7.45$ ,  $SE=0.10$  vs.  $M=7.55$ ,  $SE=0.10$  vs.  $M=7.58$ ,  $SE=0.10$ ). Results were robust to the inclusion of covariates, including AI attitudes, MMM preferences, and participant demographics. (*Supplementary Tables 4 and 5*).

Table 2: Means and standard errors of the dependent variables

| Dependent<br><i>M (SE)</i>                                | variable | Radiologist<br>only<br>n=413 | Radiologist-AI<br>agreement<br>n=401 | Radiologist<br>overcalls AI<br>n=408 | Radiologist<br>undercalls<br>AI<br>n=384 |
|-----------------------------------------------------------|----------|------------------------------|--------------------------------------|--------------------------------------|------------------------------------------|
| Agreement<br>radiologist's<br>recommendation <sup>a</sup> | with     | 4.57 (0.06)                  | 4.55 (0.07)                          | 4.63 (0.06)                          | 4.01 (0.07)                              |
| Likelihood<br>recommend radiologist <sup>a</sup>          | to       | 7.24 (0.12)                  | 7.40 (0.12)                          | 7.43 (0.11)                          | 6.47 (0.13)                              |
| Rating of radiologist <sup>a</sup>                        |          | 7.45 (0.10)                  | 7.55 (0.10)                          | 7.58 (0.10)                          | 6.82 (0.11)                              |

**Notes:**

<sup>a</sup>The radiologist-undercalls-AI group significantly differed from the three other groups on this dependent variable ( $P<.001$ ).

Agreement with the radiologist's recommendation measured from 1 = strongly disagree to 6 = strongly agree.

Likelihood to recommend radiologist measured from 1 = definitely would not recommend to 10 = definitely would recommend.

Rating of radiologist measured from 1 = worst provider possible to 10 = best provider possible.

## Moderation by MMM preferences

MMM scores had a positive effect on participants' likelihood to recommend the radiologist in the radiologist-overcalls-AI condition ( $b=0.71$ ,  $SE=0.09$ ,  $P<.001$ ). Relative to neutrals, maximizers in the radiologist-overcalls-AI condition were more likely to recommend the radiologist ( $b=1.49$ ,  $SE=0.26$ ,  $P<.001$ ).

MMM scores also had a positive effect on participants' ratings of the radiologist in the radiologist-AI agreement ( $b=0.22$ ,  $SE=0.08$ ,  $P=.005$ ) and the radiologist-overcalls-AI ( $b=0.61$ ,  $SE=0.07$ ,  $P<.001$ ) conditions. Relative to neutrals, minimizers in the radiologist-overcalls-AI condition gave lower radiologist ratings ( $b=-0.63$ ,  $SE=0.28$ ,  $P=.03$ ) whereas maximizers gave higher ratings ( $b=1.28$ ,  $SE=0.22$ ,  $P<.001$ ). Additionally, relative to neutrals, maximizers in the radiologist-undercalls-AI condition gave lower radiologist ratings ( $b=-0.49$ ,  $SE=0.23$ ,  $P=.03$ ) (*Supplementary Figures 2 and 3*).

## DISCUSSION

### Principal Results

In this study, we found that participants penalized radiologists who undercalled AI in favor of less testing but not radiologists who overcalled AI in favor of more testing. Maximizers especially disagreed with radiologists who undercalled AI. However, whereas maximizers agreed with

radiologists who overcalled AI, minimizers strongly disagreed. This suggests people's confidence in the expertise of radiologists who use AI may depend in part on whether their general preferences to maximize or minimize align with the provided recommendations.

Prior research has shown that patients demonstrate algorithm aversion and superior trust in their human clinicians versus AI [8, 29-31]. Our research suggests this may not always be the case. When AI and human clinicians are at odds, patients may trust AI over the clinician when AI detects something more than the clinician and recommends more aggressive treatment.

Although radiologists were penalized for undercalling AI, they were not penalized for overcalling AI. This overall attitude toward the radiologist who overcalled AI masked significant variability based on participants' MMM preferences. Whereas minimizers penalized the radiologist, maximizers rewarded the radiologist. The averaging of these competing attitudes resulted in a level of agreement that did not significantly differ from that in the radiologist-only condition. Thus, radiologists who overcall AI may in fact face significantly lowered trust from patients who have general minimizing preferences.

We did not observe a similar aggregate averaging effect in the radiologist-undercalls-AI condition. This asymmetry may be explained by the perceptions of minimizers. Whereas maximizers demonstrated stable preferences for pursuing more aggressive treatment [28], regardless of whether the radiologist or AI recommended it, minimizers penalized the radiologist who overcalled AI but did not reward with the radiologist who undercalled AI.

Minimizers typically prefer not to receive treatment unless considered essential [28, 32]. However, prior research has shown minimizers are more responsive to information than maximizers and will change their attitudes in response to evidence [18]. Thus, one theory for this asymmetry is that minimizers perceived AI as more accurate than the radiologist and therefore deferred to the AI's recommendation both when AI recommended more and less testing. This suggests patients, like clinicians, may be prone to automation bias when medical AI is involved [33,34]. Another theory is that the emotional salience of the AI's cancer results softened minimizers' attitudes toward more testing. Importantly, patients demonstrated higher agreement in the radiologist-only and radiologist-AI agreement conditions, despite the radiologist recommending less testing just like in the radiologist-undercalls-AI condition. This suggests participants were opposed to the radiologist's recommendation only when in contrast with the more aggressive AI recommendation. Research suggests the emotional salience of a cancer diagnosis, particularly fear of cancer, influences patients' decision making [35-37]. More research is needed to test these theories, including examining patients' general and context-specific perceptions of the accuracy of AI compared to that of clinicians [29].

Radiologists' confirmatory "second opinion" of AI did not boost participants' agreement with them. This finding adds a new dimension to literature examining patient perceptions of human- versus model-assisted decisions. In one study, patients' trust in AI increased when AI, acting as a second opinion, confirmed the human clinician's diagnosis, but decreased when the AI was disconfirming [30]. In another study, participants devalued clinicians who consulted a computer-based diagnostic aid for a second opinion, but not clinicians who consulted a human expert [38]. In this study, we found radiologists who agreed with the AI were neither rewarded nor penalized compared to the radiologist alone.

Research shows the human-AI team augments the diagnostic performance of the radiologist [39, 40]. However, we show that the human-AI team may not benefit, and could even hurt, the radiologist in

terms of patient agreement or satisfaction. This study demonstrated that even small discrepancies (i.e., 2mm) between AI and radiologist interpretations can elicit strong patient reactions. We anticipate larger discrepancies would result in more pronounced negative reactions toward clinicians. As a result, clinicians may feel compelled to overcall or agree with AI to avoid negative patient reactions, as well as potential legal consequences [41, 42]. This may lead to incorrect clinician decisions [43], as well as unnecessary testing, biopsies, and procedures. More research is needed to probe clinicians' attitudes toward the consequences of using AI.

Future research is also needed to better understand patient attitudes toward the human-AI team in various clinical contexts. Research should also interrogate what and how to communicate AI results with patients, especially when the human-AI team disagrees. Offering two recommendations to patients may eschew clinicians' professional agency and burden patients with a choice between following clinician or AI expertise [44]. However, there are ethical considerations to smoothing over discrepancies with a single recommendation that sides with either the clinician or the AI. If both recommendations are medically reasonable, a single clinician recommendation could be seen as too paternalistic [45] and may go against patient preference [46].

## Limitations

This study has several limitations. First, we intentionally recruited a diverse sample of U.S. adults to receive the broadest range of participants; however, this sample is not reflective of the patients who typically would receive LDCT scans. Future research should consider replicating this study with a narrowed sample of participants, particularly smokers, or real patients undergoing cancer screening. For example, we anticipate participants with higher pre-test probability for lung cancer would find a radiologist undercalling AI particularly difficult to accept in this scenario.

Second, we instructed participants that they were eligible to receive an LDCT scan because of "personal risk factors" but did not specify further. Risk factors could range from genetic to environmental to behavioral, and how we develop these risk factors influences our acceptance of test results [47, 48]. If participants had different mental models for how they would be eligible for lung cancer screening, these differences could have impacted their responses to the AI and radiologist's recommendations.

Third, although studies show patients want to know and consent to when AI is used [25, 49], we do not yet have a template for how healthcare systems currently, will in the future, or should report AI use to patients. In our study, the EHR results and oral follow-up provided both the AI and radiologist's interpretations and recommendations. A different presentation of AI's involvement, such as AI's interpretation but not a separate recommendation, or simply that AI was consulted but not its interpretation, could elicit different patient perceptions of their radiologist. We also do not provide measures of accuracy for either AI or radiologist. Although radiologists' accuracy is not typically disclosed to patients who receive imaging, it is possible that patients may inquire about the AI's accuracy, especially vis-à-vis the radiologist's. How patients understand and interpret measures of accuracy, including more advanced measures, such as sensitivity, specificity, precision, and recall, requires further investigation.

Moreover, future research is needed to develop validated instruments for measuring attitudes toward medical AI. Although there are validated scales measuring attitudes toward AI in general [50, 51], these were insufficient for our experimental context. Thus, the scale we deployed, for which we created a composite score from custom-designed questions and achieved a low Cronbach's alpha, was neither validated nor highly reliable, a limitation of our study. As a result, we may not have

adequately controlled for confounding medical AI attitudes, weakening our overall findings.

Last, the radiologist-patient relationship differs from the classic clinician-patient relationship, as patients might not choose, directly communicate with, or have a preexisting relationship with their radiologist [52, 53]. The nature of the radiologist-patient relationship may therefore not generalize to other medical specialties. Despite these limitations, exploring patients' attitudes toward cancer screening recommendations and toward the radiologist-patient relationship are understudied yet increasingly important topics in the era of medical AI [54].

## Conclusion

As radiologists begin to integrate AI into cancer detection, discrepancies within the human-AI team may influence patients' reactions toward their radiologist. People may penalize radiologists who undercall AI in lung cancer screening yet may not be more confident in radiologists who overcall or agree with AI. Patients' MMM preferences moderate this effect and accounting for patients' MMM preferences may give insight to clinicians on what and how much information to give about AI's role in their decision making. Our findings highlight the complexity of the patient-AI-clinician relationship and have implications for clinical practice, communication, and shared decision-making. Future research is needed to determine how radiologists should communicate AI discrepancies to patients in a way that builds trust and maintains their relevance.

## ACKNOWLEDGMENTS

*Conflicts of Interest*  
The authors do not have conflicts of interest related to this study.

*Funding*  
This study was funded by a pilot grant awarded by the University of Michigan Center for Bioethics and Social Sciences in Medicine.

*Abbreviations*

|      |                              |                       |
|------|------------------------------|-----------------------|
| AI:  | Artificial                   | intelligence          |
| MMM: | Medical                      | maximizing-minimizing |
| CT:  | Computerized                 | tomography            |
| PET: | Positron emission tomography |                       |

*Data Availability*  
The dataset generated for this study is available from the corresponding author on request.

*Authors' Contributions*  
FM, LSO, and BJZ conceptualized the study design. FM conducted data collection and formal analysis. FM, LSO, and BJZ drafted and edited the manuscript. All authors read and approved of the final manuscript.

## REFERENCES

1. Wang T-W, Hong J-S, Chiu H-Y, Chao H-S, Chen Y-M, Wu Y-T. Standalone deep learning versus experts for diagnosis lung cancer on chest computed tomography: a systematic review. *European Radiology*. 2024;1-11. PMID: 38777902
2. Gierada DS, Pinsky P, Nath H, Chiles C, Duan F, Aberle DR. Projected outcomes using different nodule sizes to define a positive CT lung cancer screening examination. *Journal of the National Cancer Institute*. 2014;106(11):dju284. PMID: 25326638
3. Tam M, Dyer T, Dissez G, Morgan TN, Hughes M, Illes J, et al. Augmenting lung cancer diagnosis on chest radiographs: Positioning artificial intelligence to improve radiologist performance. *Clinical Radiology*. 2021;76(8):607-14. PMID: 33993997
4. Rubin DL. Artificial intelligence in imaging: the radiologist's role. *Journal of the American College of Radiology*. 2019;16(9):1309-17. PMID: 31492409
5. Mikhael PG, Wohlwend J, Yala A, Karstens L, Xiang J, Takigami AK, et al. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. *Journal of Clinical Oncology*. 2023;41(12):2191-200. PMID: 36634294
6. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR Data Science Institute artificial intelligence survey. *Journal of the American College of Radiology*. 2021;18(8):1153-9. PMID: 33891859
7. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. *Journal of Consumer Research*. 2019;46(4):629-50. DOI: 10.1093/jcr/ucz013
8. Cadario R, Longoni C, Morewedge CK. Understanding, explaining, and utilizing medical artificial intelligence. *Nature human behaviour*. 2021;5(12):1636-42. PMID: 34183800
9. Promberger M, Baron J. Do patients trust computers? *Journal of Behavioral Decision Making*. 2006;19(5):455-68. DOI: 10.1002/bdm.542
10. Gaczek P, Pozharliev R, Leszczyński G, Zieliński M. Overcoming consumer resistance to AI in general health care. *Journal of Interactive Marketing*. 2023;58(2-3):321-38. DOI: 10.1177/10949968221151061
11. Food U, Administration D. Good machine learning practice for medical device development: guiding principles. The US Food and Drug Administration: Washington, DC, USA. 2021.
12. Birkhäuser J, Gaab J, Kossowsky J, Hasler S, Krummenacher P, Werner C, et al. Trust in the health care professional and health outcome: A meta-analysis. *PloS one*. 2017;12(2):e0170988. PMID: 28170443
13. Lee Y-Y, Lin JL. The effects of trust in physician on self-efficacy, adherence and diabetes outcomes. *Social science & medicine*. 2009;68(6):1060-8. PMID: 19162386
14. Thom DH, Hall MA, Pawlson LG. Measuring patients' trust in physicians when assessing quality of care. *Health affairs*. 2004;23(4):124-32. PMID: 15318572
15. Zondag AG, Rozestraten R, Grimmelikhuijsen SG, Jongsma KR, van Solinge WW, Bots ML, et al. The Effect of Artificial Intelligence on Patient-Physician Trust: Cross-Sectional Vignette Study. *Journal of Medical Internet Research*. 2024;26:e50853. PMID: 38805702
16. Derevianko A, Pizzoli SFM, Pesapane F, Rotili A, Monzani D, Grasso R, et al. The use of artificial intelligence (AI) in the radiology field: What is the state of doctor-patient communication in cancer diagnosis? *Cancers*. 2023;15(2):470. PMID: 36672417
17. Scherer LD, Zikmund-Fisher BJ. Eliciting medical maximizing-minimizing preferences with a single question: development and validation of the MM1. *Medical Decision Making*. 2020;40(4):545-50. PMID: 32522094
18. Scherer LD, Kullgren JT, Caverly T, Scherer AM, Shaffer VA, Fagerlin A, et al. Medical maximizing-minimizing preferences predict responses to information about prostate-specific antigen screening. *Medical decision making*. 2018;38(6):708-18. PMID: 29916784
19. Evron JM, Reyes-Gastelum D, Banerjee M, Scherer LD, Wallner LP, Hamilton AS, et al. Role

- of patient maximizing-minimizing preferences in thyroid cancer surveillance. *Journal of Clinical Oncology*. 2019;37(32):3042-9. PMID: 31573822
20. Kang SK, Scherer LD, Megibow AJ, Higuera LJ, Kim N, Braithwaite RS, et al. A randomized study of patient risk perception for incidental renal findings on diagnostic imaging tests. *American Journal of Roentgenology*. 2018;210(2):369-75. PMID: 29140116
21. Vordenberg SE, Zikmund-Fisher BJ. Characteristics of older adults predict concern about stopping medications. *Journal of the American Pharmacists Association*. 2020;60(6):773-80. PMID: 32115393
22. Scherer LD, Shaffer VA, Caverly T, DeWitt J, Zikmund-Fisher BJ. Medical maximizing-minimizing predicts patient preferences for high-and low-benefit care. *Medical Decision Making*. 2020;40(1):72-80. PMID: 31975657
23. Smith KT, Monti D, Mir N, Peters E, Tipirneni R, Politi MC. Access is necessary but not sufficient: factors influencing delay and avoidance of health care services. *MDM Policy & Practice*. 2018;3(1):2381468318760298. PMID: 30288438
24. AHRQ. CAHPS Patient Experience Surveys and Guidance. AHRQ: AHRQ; 2023 [updated September 2023; cited 2024 August 13, 2024]; Available from: <https://www.ahrq.gov/cahps/surveys-guidance/index.html>.
25. Khullar D, Casalino LP, Qian Y, Lu Y, Krumholz HM, Aneja S. Perspectives of patients about artificial intelligence in health care. *JAMA network open*. 2022;5(5):e2210309-e. PMID: 35507346
26. Fritsch SJ, Blankenheim A, Wahl A, Hetfeld P, Maassen O, Deffge S, et al. Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients. *Digital health*. 2022;8:20552076221116772. PMID: 35983102
27. Dossett LA, Mott NM, Bredbeck BC, Wang T, Jobin CT, Hughes TM, et al. Using tailored messages to target overuse of low-value breast cancer care in older women. *Journal of Surgical Research*. 2022;270:503-12. PMID: 34801801
28. Mott N, Wang T, Miller J, Berlin NL, Hawley S, Jagsi R, et al. Medical maximizing-minimizing preferences in relation to low-value services for older women with hormone receptor-positive breast cancer: a qualitative study. *Annals of surgical oncology*. 2021;28:941-9. PMID: 32720038
29. Robertson C, Woods A, Bergstrand K, Findley J, Balser C, Slepian MJ. Diverse patients' attitudes towards Artificial Intelligence (AI) in diagnosis. *PLOS Digital Health*. 2023;2(5):e0000237. PMID: 37205713
30. Juravle G, Boudouraki A, Terziyska M, Rezlescu C. Trust in artificial intelligence for medical diagnoses. *Progress in brain research*. 2020;253:263-82. PMID: 32771128
31. York T, Jenney H, Jones G. Clinician and computer: a study on patient perceptions of artificial intelligence in skeletal radiography. *BMJ Health & Care Informatics*. 2020;27(3). PMID: 33187956
32. Scherer LD, Caverly TJ, Burke J, Zikmund-Fisher BJ, Kullgren JT, Steinley D, et al. Development of the medical maximizer-minimizer scale. *Health Psychology*. 2016;35(11):1276. PMID: 27617512
33. Jabbour S, Fouhey D, Shepard S, Valley TS, Kazerooni EA, Banovic N, et al. Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. *Jama*. 2023;330(23):2275-84. PMID: 38112814
34. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*. 2012;19(1):121-7. PMID: 21685142
35. Mazzocco K, Masiero M, Carriero MC, Pravettoni G. The role of emotions in cancer patients' decision-making. *Ecancermedicalscience*. 2019;13:914. PMID: 31123497
36. Nold RJ, Beamer RL, Helmer SD, McBoyle MF. Factors influencing a woman's choice to undergo breast-conserving surgery versus modified radical mastectomy. *The American journal of*

surgery. 2000;180(6):413-8. PMID: 11182389

37. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Risky feelings: why a 6% risk of cancer does not always feel like 6%. *Patient education and counseling*. 2010;81:S87-S93. PMID: 20739135

38. Probst CA, Shaffer VA, Lambdin C, Arkes HR, Medow MA. Ratings of physicians relying on experts versus physicians relying on decision aids. In *Proceedings: 4<sup>th</sup> Annual Symposium: Graduate Research and Scholarly Projects*. Wichita, KS: Wichita State University. 2008; 91-92.

39. Li D, Pehrson LM, Lauridsen CA, Tøttrup L, Fraccaro M, Elliott D, et al. The added effect of artificial intelligence on physicians' performance in detecting thoracic pathologies on CT and chest X-ray: A systematic review. *Diagnostics*. 2021;11(12):2206. PMID: 34943442

40. Bennani S, Regnard N-E, Ventre J, Lassalle L, Nguyen T, Ducarouge A, et al. Using AI to improve Radiologist performance in detection of abnormalities on chest Radiographs. *Radiology*. 2023;309(3):e230860. PMID: 38085079

41. Lawton T, Morgan P, Porter Z, Hickey S, Cunningham A, Hughes N, et al. Clinicians risk becoming 'liability sinks' for artificial intelligence. *Future Healthcare Journal*. 2024;11(1). PMID: 38646041

42. Smith H, Fotheringham K. Artificial intelligence in clinical decision-making: rethinking liability. *Medical Law International*. 2020;20(2):131-54. DOI: 10.1177/0968533220945766

43. Bernstein MH, Atalay MK, Dibble EH, Maxwell AW, Karam AR, Agarwal S, 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. *European radiology*. 2023;33(11):8263-9. PMID: 37266657

44. Kilbride MK, Joffe S. The new age of patient autonomy: implications for the patient-physician relationship. *JAMA*. 2018;320(19):1973-4. PMID: 30326026

45. Quill TE, Brody H. Physician recommendations and patient autonomy: finding a balance between physician power and patient choice. *Annals of internal medicine*. 1996;125(9):763-9. PMID: 8929011

46. Gurmankin AD, Baron J, Hershey JC, Ubel PA. The role of physicians' recommendations in medical treatment decisions. *Medical Decision Making*. 2002;22(3):262-71. PMID: 12058783

47. Bostrom A. Mental models of risk. *Oxford Research Encyclopedia of Communication*. 2017.

48. Holtrop JS, Scherer LD, Matlock DD, Glasgow RE, Green LA. The importance of mental models in implementation science. *Frontiers in Public Health*. 2021;9:680316. PMID: 34295871

49. Moy S, Irannejad M, Manning SJ, Farahani M, Ahmed Y, Gao E, et al. Patient Perspectives on the Use of Artificial Intelligence in Health Care: A Scoping Review. *Journal of Patient-Centered Research and Reviews*. 2024;11(1):51. PMID: 38596349

50. Schepman A, Rodway P. Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in human behavior reports*. 2020;1:100014. PMID: 34235291

51. Grassini S. Development and validation of the AI attitude scale (AIAS-4): a brief measure of general attitude toward artificial intelligence. *Frontiers in psychology*. 2023;14:1191628. PMID: 37554139

52. Koney N, Roudenko A, Ro M, Bahl S, Kagen A. Patients want to meet with imaging experts. *Journal of the American College of Radiology*. 2016;13(4):465-70. PMID: 26857485

53. Gunn AJ, Mangano MD, Choy G, Sahani DV. Rethinking the role of the radiologist: enhancing visibility through both traditional and nontraditional reporting practices. *Radiographics*. 2015;35(2):416-23. PMID: 25763726

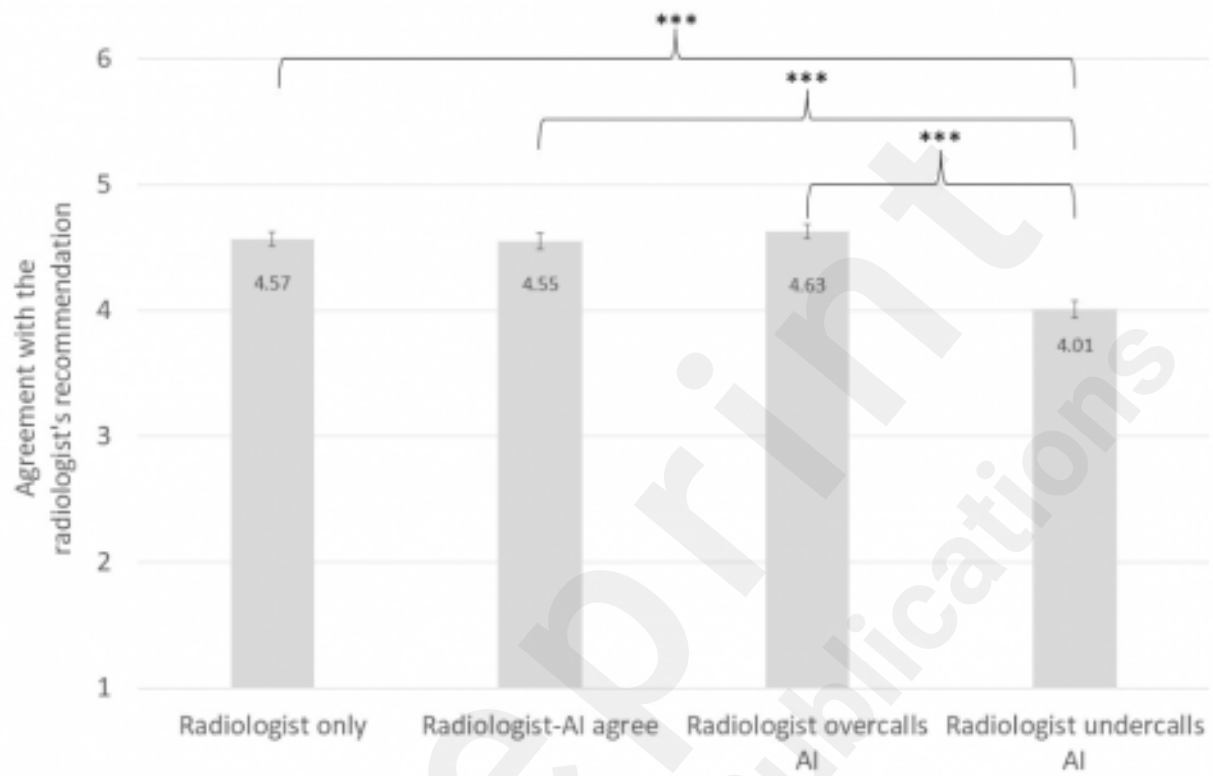
54. Kitts AB. Patient perspectives on artificial intelligence in radiology. *Journal of the American College of Radiology*. 2023;20(9):863-7. PMID: 3745601

## Supplementary Files

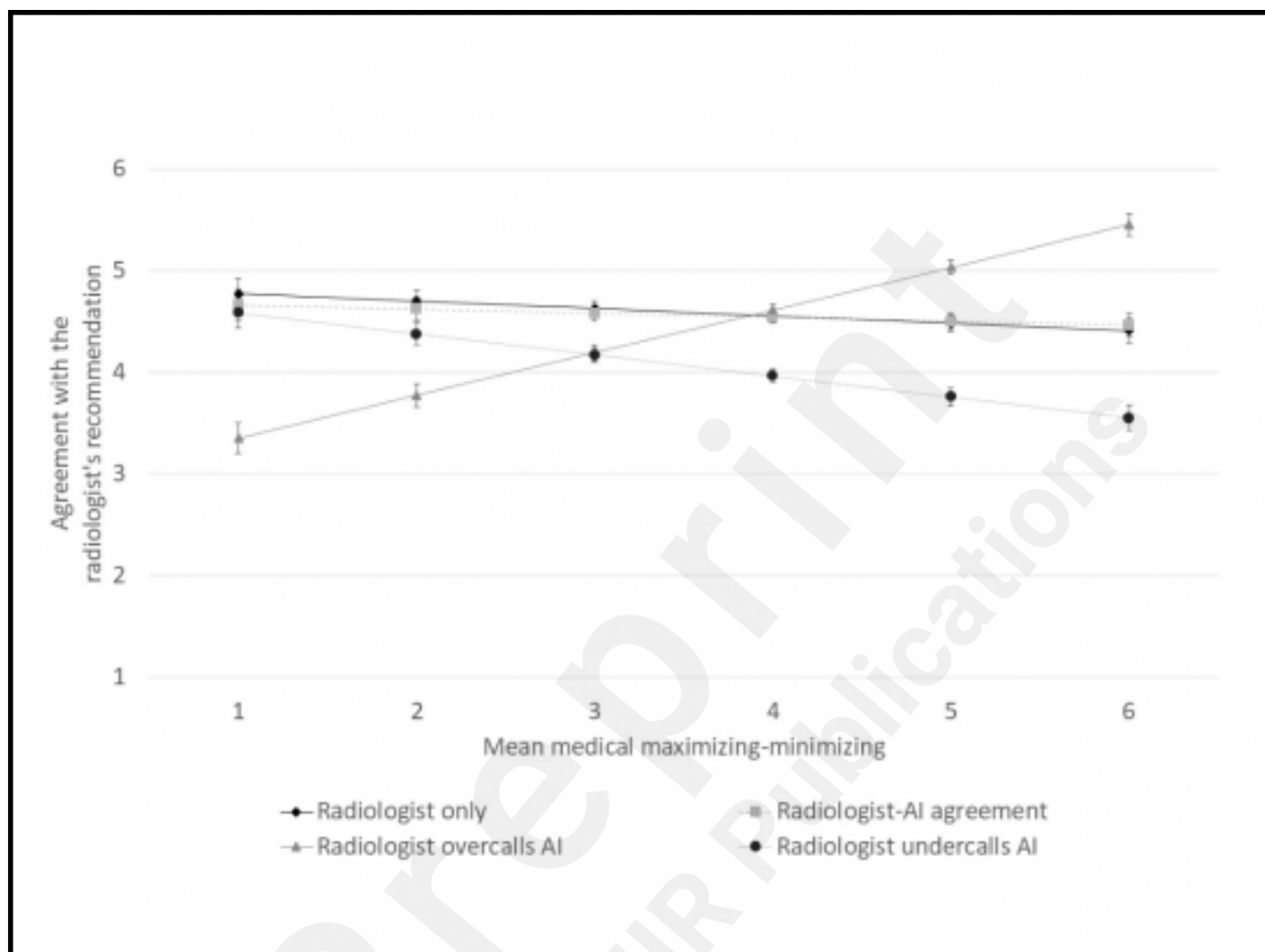
## Figures

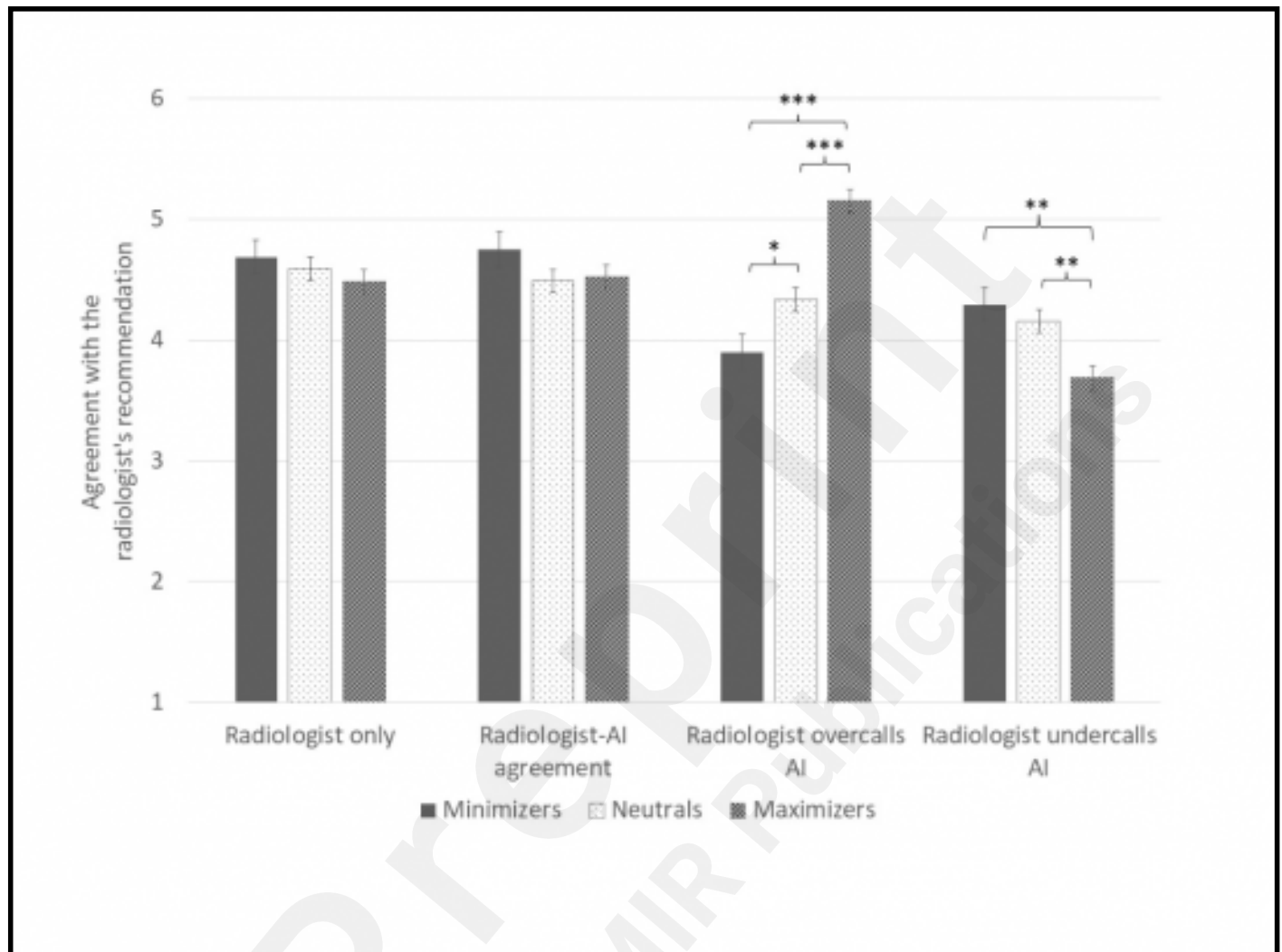
Hypothetical results of the LDCT scan, formatted to align with common EHR systems, received by participants, by condition.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p style="text-align: center;"><b>Radiologist only</b></p> <p><u>IMPRESSION:</u></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p>Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><u>RECOMMENDATION:</u></p> <p>Repeat low-dose CT in 6 months to ensure no interval growth.</p>                                                                                                                                                                                                                                                                                                                    | <p style="text-align: center;"><b>Radiologist-AI agreement</b></p> <p><u>IMPRESSION:</u></p> <p><b>LungDetect AI:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p>Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><b>Radiologist:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p>Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><u>RECOMMENDATION:</u></p> <p>Repeat low-dose CT in 6 months to ensure no interval growth, based on AI and radiologist agreement.</p>                                                            |
| <p style="text-align: center;"><b>Radiologist overcalls AI</b></p> <p><u>IMPRESSION:</u></p> <p><b>LungDetect AI:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p>Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><b>Radiologist:</b></p> <p>1. There is a 9-millimeter solid nodule in the left lung.</p> <p>Lung-RADS 4A: Suspicious. <b>Medium risk of malignancy (cancer) 5-15%</b></p> <p><u>RECOMMENDATION:</u></p> <p><b>LungDetect AI:</b><br/>Repeat low-dose CT in 6 months to ensure no interval growth.</p> <p><b>Radiologist:</b><br/>Undergo additional imaging immediately.</p> | <p style="text-align: center;"><b>Radiologist undercalls AI</b></p> <p><u>IMPRESSION:</u></p> <p><b>LungDetect AI:</b></p> <p>1. There is a 9-millimeter solid nodule in the left lung.</p> <p>Lung-RADS 4A: Suspicious. <b>Medium risk of malignancy (cancer) 5-15%</b></p> <p><b>Radiologist:</b></p> <p>1. There is a 7-millimeter solid nodule in the left lung.</p> <p>Lung-RADS 3: Probably benign. <b>Low risk of malignancy (cancer) 1-2%</b></p> <p><u>RECOMMENDATION:</u></p> <p><b>LungDetect AI:</b><br/>Undergo additional imaging immediately.</p> <p><b>Radiologist:</b><br/>Repeat low-dose CT in 6 months to ensure no interval growth.</p> |



Effect of participants' medical maximizing-minimizing preferences on agreement with the radiologist's recommendation, by condition. Standard error bars depicted.





## Multimedia Appendixes

Supplementary material.

URL: <http://asset.jmir.pub/assets/221376fd5c644c44eb4d2c686209198c.docx>

