Effect of uncertainty-aware artificial intelligence models on human reaction time and decision-making: A randomized controlled trial

Corey Lester, Brigid Rowell, Yifan Zheng, Zoe Co, Vincent Marshall, Jin Yong Kim, Qiyuan Chen, Raed Kontar, X. Jessie Yang

Submitted to: Journal of Medical Internet Research
on: July 30, 2024

Table of Contents

Effect of uncertainty-aware artificial intelligence models on human reaction time and decision-making: A randomized controlled trial

Corey Lester[1] PharmD, PhD; Brigid Rowell[1] MA; Yifan Zheng[1] PharmD; Zoe Co[1, 2] BS; Vincent Marshall[1] MS; Jin Yong Kim[3] BSE; Qiyuan Chen[3] MSE; Raed Kontar[3] PhD; X. Jessie Yang[3] PhD

[1]Department of Clinical Pharmacy College of Pharmacy University of Michigan Ann Arbor US
[2]Department of Learning Health Sciences University of Michigan School of Medicine Ann Arbor US
[3]Department of Industrial and Operations Engineering College of Engineering University of Michigan Ann Arbor US

Corresponding Author:
Corey Lester PharmD, PhD
Department of Clinical Pharmacy
College of Pharmacy
University of Michigan
428 Church Street
Ann Arbor
US

Abstract

Background: Artificial intelligence (AI)-based clinical decision support systems are increasingly used in healthcare. Uncertainty-aware AI presents the model's confidence in its decision alongside its prediction whereas black-box AI only provides a prediction. Little is known about how this type of AI affects healthcare providers' work performance and reaction time.

Objective: To determine the effects of black-box and uncertainty-aware AI advice on pharmacist decision-making and reaction time.

Methods: Thirty licensed pharmacists participated in a crossover, randomized controlled trial. Eligible participants were randomized to either the black-box AI or uncertainty-aware AI condition in a 1:1 manner. Participants completed 100 mock verification tasks with AI help and 100 without AI help. The order of no help and AI help was randomized. Participants were exposed to correct and incorrect prescription fills, where the correct decision was to 'accept' or 'reject', respectively. AI help provided correct (79%) or incorrect (21%) advice. Reaction times, participant decision, AI advice, and AI help type were recorded for each verification. Likelihood ratio tests (LRT) compared means across the three categories of AI type for each level of AI correctness.

Results: Participants' decision-making performance and reaction times differed across the three conditions. Accurate AI recommendations resulted in the rejection of the incorrect drug 96.1% and 91.8% of the time for uncertainty-aware AI and black-box AI respectively, compared to 81.2% without AI help. Correctly dispensed medications were accepted at rates of 99.2% with black-box help, 94.1% with uncertainty-aware AI help, and 94.6% without AI help. Uncertainty-aware AI protected against bad AI advice to approve an incorrectly filled medication compared to black-box AI (83.3% vs 76.7%). When the AI recommended rejecting a correctly filled medication, pharmacists without AI help had a higher rate of correctly accepting the medication (94.6%) compared to uncertainty-aware AI help (86.2%) and black-box AI help (81.2%). Uncertainty-aware AI resulted in shorter reaction times than black-box AI and no AI help except in the scenario where "AI rejects the correct drug". Black-box AI did not lead to reduced reaction times compared to pharmacists acting alone.

Conclusions: Pharmacists' performance and reaction times varied by AI type and AI accuracy. Overall, uncertainty-aware AI resulted in faster decision-making and acted as a safeguard against bad AI advice to approve a misfilled medication. Conversely, black-box AI had the longest reaction times, and user performance degraded in the presence of bad AI advice. However, uncertainty-aware AI could result in unnecessary double-checks, but it is preferred over false negative advice, where patients receive the wrong medication. These results highlight the importance of well-designed AI that addresses users' needs, enhances performance, and avoids overreliance on AI.

(JMIR Preprints 30/07/2024:64902)
DOI: https://doi.org/10.2196/preprints.64902

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 v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Effect of uncertainty-aware artificial intelligence models on human reaction time and decision-making: A randomized controlled trial

Corey Lester, PharmD, PhD. Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, Ann Arbor, MI, United States
Brigid Rowell, MA. Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, Ann Arbor, MI, United States
Zoe Co, BS. Department of Learning Health Sciences, University of Michigan School of Medicine, Ann Arbor, MI, United States, Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, Ann Arbor, MI, United States
Vincent D Marshall, MS. Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, Ann Arbor, MI, United States
Jin Yong Kim, BSE. Department of Industrial and Operations Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States
Qiyuan Chen, MSE. Department of Industrial and Operations Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States
Raed Al Kontar, PhD. Department of Industrial and Operations Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States
X. Jessie Yang, PhD. Department of Industrial and Operations Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States

Corresponding author: Corey Lester, PharmD, PhD., Department of Clinical Pharmacy, College of Pharmacy,University of Michigan, Ann Arbor, MI 48109, 734-647-8849, lesterca@umich.edu

Keywords: artificial intelligence, human-computer interaction, decision-making, human factors

Abstract:

Background: Artificial intelligence (AI)-based clinical decision support systems are increasingly used in healthcare. Uncertainty-aware AI presents the model's confidence in its decision alongside its prediction whereas black-box AI only provides a prediction. Little is known about how this type of AI affects healthcare providers' work performance and reaction time.

Objective: To determine the effects of black-box and uncertainty-aware AI advice on pharmacist decision-making and reaction time.

Methods: Thirty licensed pharmacists participated in a crossover, randomized controlled trial. Eligible participants were randomized to either the black-box AI or uncertainty-aware AI condition in a 1:1 manner. Participants completed 100 mock verification tasks with AI help and 100 without AI help. The order of no help and AI help was randomized. Participants were exposed to correct and incorrect prescription fills, where the correct decision was to 'accept' or 'reject', respectively. AI help provided correct (79%) or incorrect (21%) advice. Reaction times, participant decision, AI advice, and AI help type were recorded for each verification. Likelihood ratio tests (LRT) compared means across the three categories of AI type for each level of AI correctness.

Results: Participants' decision-making performance and reaction times differed across the three conditions. Accurate AI recommendations resulted in the rejection of the incorrect drug 96.1% and 91.8% of the time for uncertainty-aware AI and black-box AI respectively, compared to 81.2% without AI help. Correctly dispensed medications were accepted at rates of 99.2% with black-box help, 94.1% with uncertainty-aware AI help, and 94.6% without AI help. Uncertainty-aware AI protected against bad AI advice to approve an incorrectly filled medication compared to black-box AI (83.3% vs 76.7%). When the AI recommended rejecting a correctly filled medication, pharmacists without AI help had a higher rate of correctly accepting the medication (94.6%) compared to uncertainty-aware AI help (86.2%) and black-box AI help (81.2%). Uncertainty-aware AI resulted in shorter reaction times than black-box AI and no AI help except in the scenario where "AI rejects the correct drug". Black-box AI did not lead to reduced reaction times compared to pharmacists acting alone.

Conclusions: Pharmacists' performance and reaction times varied by AI type and AI accuracy. Overall, uncertainty-aware AI resulted in faster decision-making and acted as a safeguard against bad AI advice to approve a misfilled medication. Conversely, black-box AI had the longest reaction times, and user performance degraded in the presence of bad AI advice. However, uncertainty-aware AI could result in unnecessary double-checks, but it is preferred over false negative advice, where patients receive the wrong medication. These results highlight the importance of well-designed AI that addresses users' needs, enhances performance, and avoids overreliance on AI.

Introduction:

Artificial intelligence (AI) is becoming increasingly prevalent in healthcare with a wide range of applications such as drug development,[1] computer-aided diagnosis and detection,[2,3] and clinical decision-making.[4] In particular, AI-based clinical decision support systems (CDSS) can improve medication safety and reduce medication errors. CDSS have been shown to significantly improve medication use and safety in areas such as drug-drug interactions, inappropriate prescribing in the elderly and pregnant women, over- and underprescribing, patient counseling[5,6], and patient outcomes (e.g., increased medication adherence, lower blood pressure, decreased adverse events).[6]

Recent advances in AI modeling such as deep neural networks have resulted in CDSS that are “black-box” systems; black-box outputs do not provide insight into the model’s decision-making process or confidence in its decision.[7] To address the calls for increased transparency in medical AI predictions,[8–10] developers have started employing uncertainty-aware AI models. Uncertainty-aware AI models present the model’s uncertainty, or confidence in its decision, alongside its prediction[11] thus providing a metric for the user to assess the AI’s reliability.[12] CDSS reliability is an essential component of human evaluation of AI’s trustworthiness which determines the user’s acceptability of a technology.[7] While uncertainty-aware AI models increase transparency, additional knowledge is required to interpret the findings which may initially confuse users, leading to increased cognitive effort and degraded decision-making.[13]

In addition to transparency, automation bias and aversion must be considered when developing CDSS. Automation bias occurs when users forsake their own expertise in favor of the AI’s advice.[12,14] The overreliance caused by automation bias may result in users missing AI-generated errors. Automation aversion causes users to rapidly and persistently lose trust due to an AI-generated error even when the AI’s overall performance exceeds humans.[12] Consequently, human errors are missed. An ideal human-centered AI tool should generate clinically valid decisions while fostering trust and avoiding overreliance on AI.

An unexplored avenue for AI-based CDSS is the medication verification process. Medication verification is a vital yet time-consuming visual check to ensure the contents of a filled medication vial match the prescribed medication. Despite pharmacists’ careful medication verification process, dispensing errors occur in 1.5% of all prescriptions.[15] Vigilance, or the ability to maintain focus and alertness over long periods, is essential for repeated tasks such as medication verification.[16] Pharmacists must remain alert and cognitively engaged;[17] yet, human vigilance wanes over time. Waning vigilance may be due to fatigue, cognitive overload, or the mundanity of repetitive tasks.[18,19] This is especially concerning in community pharmacies due to the increasing prescription volumes. In 2022 the average community pharmacy dispensed approximately 1,215 prescriptions per week.[20] The increase in prescription volumes necessitates additional effort and time from pharmacists who spend between 30-48% of their time verifying medications.[21,22]

Our team developed an AI prototype to assist pharmacists with the medication verification task with the goal of reducing dispensing errors, improving patient safety, and decreasing pharmacists' workload. The development of our AI tool using user-centered design principles is described in Zheng et al.[23] The purpose of this paper is to determine the effects of black-box and uncertainty-aware AI advice on pharmacist decision-making and reaction time.

Methods:

We previously developed two AI conditions to assist pharmacists with making decisions when dispensing medications. The "black-box" condition does not provide the user with insight into the AI's certainty or predicted probability of a correct decision. The "uncertainty-aware" condition provides the user with an estimate of certainty from the model predictions. We test these conditions in an experimental study using mock verification tasks with pharmacists.

Trial design

A crossover, randomized controlled trial was conducted from January 2023 to May 2023 with licensed pharmacists in the United States. Eligible participants were randomized to either the black-box AI or uncertainty-aware AI condition in a 1:1 manner. All participants completed 100 mock verifications with AI help and 100 mock verifications without AI help. The order of no help and AI help was randomized. The University of Michigan Institutional Review Board determined this research met the criteria for Exemption #3 and was exempt from IRB oversight. All participants signed a prospective agreement prior to any research activities.

Participants

Recruitment emails describing the study were sent to pharmacists through the Minnesota Pharmacy Practice-Based Research Network listserv and the University of Michigan College of Pharmacy Preceptor Network listserv. Interested individuals contacted the study team directly and completed a screening phone call. Eligible participants were licensed pharmacists in the United States who were at least 18 years old and had access to a laptop or desktop computer with a webcam. Pharmacists were excluded if they: 1) require assistive technology to use the computer, 2) wear eyeglasses with more than one power, 3) have uncorrected cataracts, intraocular implants, glaucoma, or permanently dilated pupils, or 4) have eye movement or alignment abnormalities (e.g., lazy eye, strabismus, nystagmus).

AI input

We previously obtained a dataset of 432,974 images from a mail-order pharmacy that fills and ships prescriptions to all 50 US States.[24] The data contain 1 year's worth of images of oral medications (i.e., tablets and capsules) inside a prescription vial filled by a robot. The images were taken as the final step of an automated system using a robot to count pills into a vial, label the vial, take photos of the vial's contents, and cap the vial. The dataset images are aligned with a National Drug Code (NDC) label and different attributes including color, shape, size in millimeters of pills, manufacturer, tablet scoring, and imprint. The number of images for each NDC ranges from 3 to 12,105 with a median of 540. There are 12 different colors of medications labeled in these images: white (42.1%), yellow (12.3%), pink (9.1%), orange (7.1%), multi-color (5.9%), green (5.2%), red (5.2%), blue (4.8%), brown (3.8%), purple (3.1%), turquoise (0.7%), and gray (0.7%). Seven different shapes are identified in the data: round (49.6%), oval (33.4%), capsule (16.2%), hexagon-6-sided (0.4%), triangle (0.3%), trapezoid (0.1%) and pentagon-five sided (0.0%).

Interventions

The AI model in our study refers to a Bayesian neural network that predicts the dispensed pills' NDC along with the uncertainty of the predictions. It is realized by applying the random dropout technique[25] to the ResNet-34[26] convolutional neural network. Rather than simply predicting the probabilities of belonging to a specific NDC, the dropout technique enables the neural network to sample a set of possible predictions which we use to measure the uncertainty of the prediction. In our research, the model generated 50 potential probabilities for every image.

Fill accuracy is a dichotomous variable (correct fill/incorrect fill) determined by comparing the fill image to the reference image (ground truth). Matching fill and reference image pairs are labeled correct fills and mismatched pairs are labeled incorrect fills. This serves as the ground truth for the image pair. AI accuracy (good prediction/bad prediction) is a variable that indicates the AI's accuracy in predicting the fill image's NDC. A "good prediction" means the AI correctly identified the NDC whereas a "bad prediction" incorrectly predicted the NDC. Good predictions always recommend the correct user action (accept or reject) whereas bad predictions typically result in ill-advised recommendations.

In each AI condition of 100 trials, the AI accurately recommended accepting (AI Accept) the correct fill in 60 trials and incorrectly recommended rejecting (AI Reject) an incorrect fill in 16 trials. For incorrect fills, 22 trials correctly were AI Reject and 2 trials were AI Accept. Three of the AI Reject trials contained an incorrect fill coupled with a bad prediction resulting in the correct recommendation (AI Reject). Although the AI's predicted NDC was incorrect (i.e., misidentified the fill image NDC), the AI renders the correct advice to reject the mis-filled medication.

While the AI model's overall accuracy is 98.46%, it was lowered to 79% for the experiment. Participants' interaction with AI errors is a critical component of the research. Lowering the model's accuracy significantly reduced the number of AI help trials needed to display the requisite number of AI errors. In the no-help condition, 76% of the medications were correctly filled.

The NDCs, the reference images, and the filled images were gathered from the correctly and incorrectly predicted images of the Bayesian neural network model. To eliminate potential confounding variables, each reference NDC was shown no more than twice throughout the experiment. To avoid loss of variation in colors, capsules, and oddly shaped pills while excluding blurry fill images, the team members carefully selected from the model-predicted images. Each image was reviewed for accuracy by comparing a reference image (ground truth) to the fill image.

There were forty unique NDCs in the incorrect prediction file of the Bayesian neural network model. The NDCs, the reference images, and the images for cases involving bad machine predictions (AI approves the incorrect drug, and AI rejects the correct drug) were gathered from the incorrect prediction file. The NDCs and the reference images for cases involving good machine prediction for the incorrectly filled medication (AI rejects the incorrect drug) were gathered from the incorrect prediction file. The filled images were gathered from the correct prediction file.

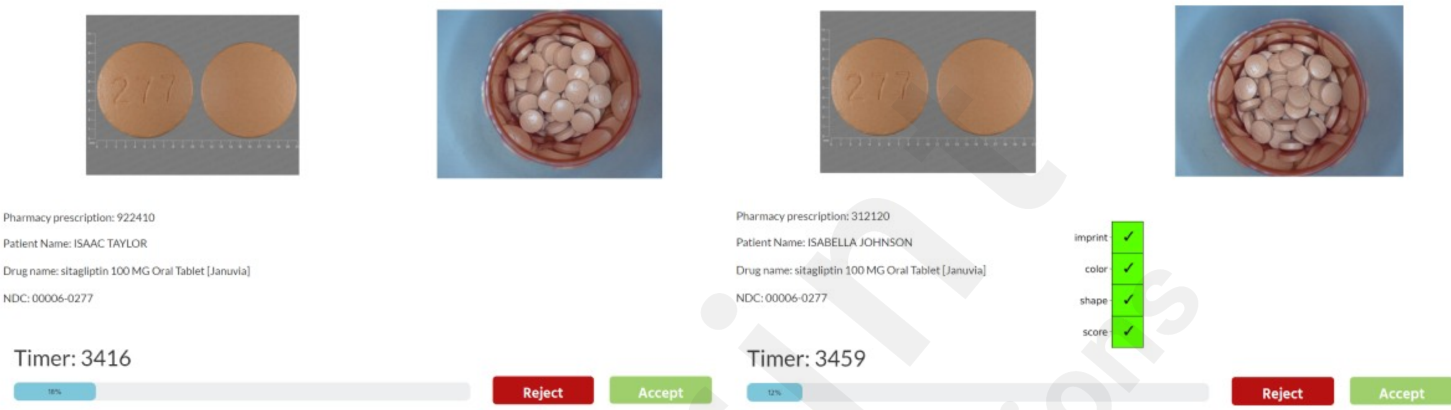
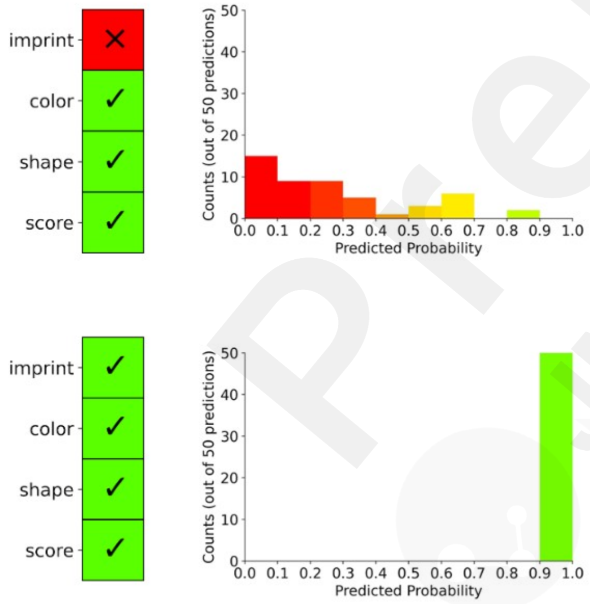


Figure 1. An example of the interface. The task was to compare the “reference image” on the left to the “fill image” on the right to determine whether to accept or reject the dispensed medication. The “reference image” is the ground truth. For half the trials there was AI help on the bottom right of the screen.



The study used Labvanced™, an online, browser-based platform, to conduct the mock verifications. A video presentation explained how to perform the mock medication verification task using the study interface. Participants were informed that the goal of the task was to determine whether an image of a filled medication bottle matched a known reference image. The video presentation also explained the AI help condition to the participant with a tailored explanation based on randomization assignment. The interface displayed an image of a filled medication, a reference image, and prescription information (Figure 1). In the AI help conditions, the interface displayed AI advice as 1) green checkboxes indicating AI matches based on characteristics of the pill (i.e., black box), or 2) a histogram with predicted probabilities (i.e., uncertainty) and the checkbox figure (Figure 2). The AI advice was created from a ResNet-34 neural network.

Figure 2. Presentations of two AI models. The left-hand graphs are “black-box AI” which displays “yes” or “no” for four pill characteristics: imprint, color, shape, and score. The right-hand graphs are “uncertainty-aware AI” which illustrates the predicted probability of each set of predictions summarized in a histogram.

Each participant completed 200 verification tasks, a series of 100 verifications with AI help, and a series of 100 verifications without AI help. Presentation of AI help and no AI was block randomized. The pharmacist's decision to approve or reject the filled medication and the time to decide were recorded for each verification. After each verification in the AI help condition, participants rated how much they trusted the AI advice on a scale of 0 to 100.

Analyses

The outcomes of accuracy and reaction time were analyzed in mixed model generalized linear models with categorical predictors. Categorical predictors included the AI type (i.e., black-box AI, uncertainty-aware AI, or no AI help), and AI correctness vs. ground truth, making the 2x2 comparison with 4 categories. Likelihood ratio tests (LRT) compared means across the three categories of AI type for each level of AI correctness. The simple effects of uncertainty-aware AI and black-box AI and user performance based on AI type were examined. Statistical comparisons were performed with R statistical software (version 4.2.2).[27]

Randomization

A random number generator in R was used to ensure unbiased study designs. Numbers from 1 to 8 were randomly generated for each participant and participants had an equal probability of being assigned to either the uncertainty-aware AI condition or the black-box AI condition at the start. An equal number of participants were sought for each of the 8 study conditions.

Results

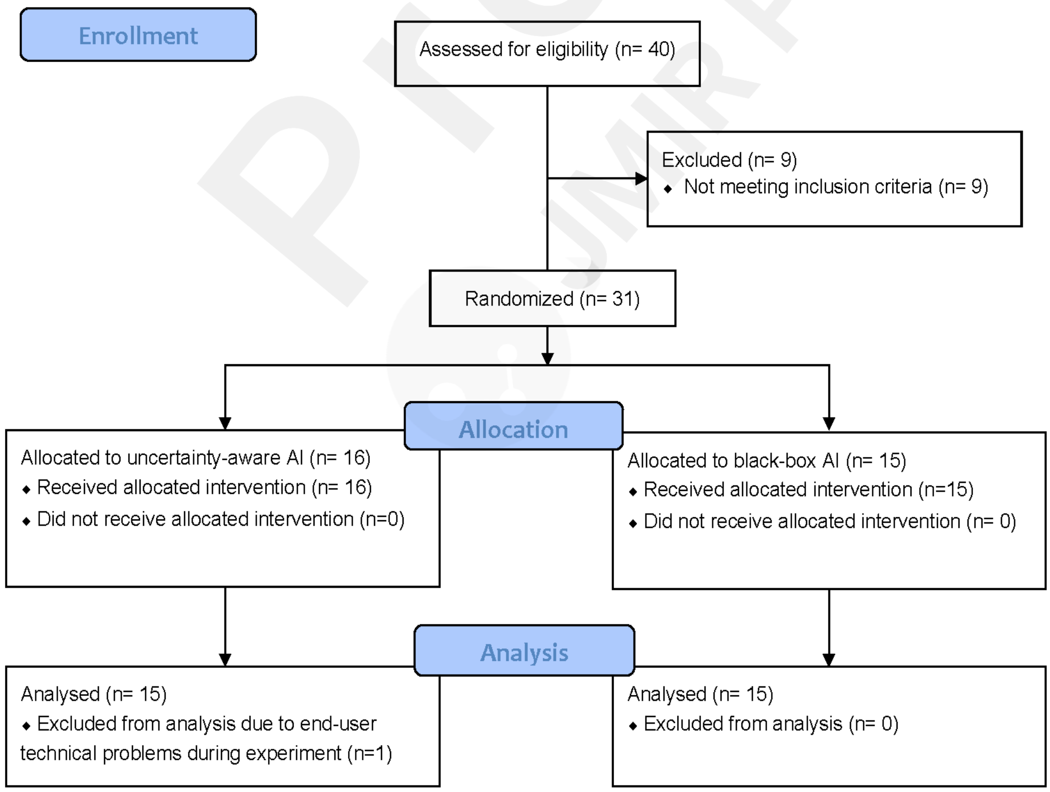


Figure 3. Participant flow of the RCT for effect of artificial intelligence models on medication dispensing in pharmacists.

Recruitment

Recruitment began in January 2023 and the final participant visit was in May 2023. The single study visit was conducted remotely using a laptop or desktop computer. The trial ended when the target enrollment was met. Forty pharmacists were assessed for eligibility. After excluding 9 pharmacists who did not meet the inclusion criteria, 31 participants were randomized for allocation in the “black-box AI” condition or the “uncertainty-aware AI” condition. Ultimately, 15 subjects in each condition completed the experiment and were included in the subsequent analysis.

	Overall	Uncertainty-Aware	Black-box	*P*	effect size	interpretation
n	30	15	15			
Age (median [IQR])	36.00 [32.25, 45.75]	35.73 (9.68)	43.07 (11.79)	0.073	0.68*	medium
Gender = male (%)	13 (43.3)	7 (46.7)	6 (40.0)	>0.999	0.07	small
Ethnicity = Not Hispanic or Latino (%)	28 (93.3)	15 (100.0)	13 (86.7)	0.464	0.27	medium
Race (%)				0.549	0.2	small
Asian	5 (16.7)	3 (20.0)	2 (13.3)			
More than one race	1 (3.3)	0 (0.0)	1 (6.7)			
White	24 (80.0)	12 (80.0)	12 (80.0)			
Practice Setting (%)				0.248	0.42	large
Community pharmacy	15 (50.0)	6 (40.0)	9 (60.0)			
Grocery store / mass merchandise pharmacy	1 (3.3)	1 (6.7)	0 (0.0)			
Hospital Pharmacy	6 (20.0)	5 (33.3)	1 (6.7)			
Other	7 (23.3)	3 (20.0)	4 (26.7)			
Specialty Pharmacy	1 (3.3)	0 (0.0)	1 (6.7)			
Years Worked (%)				0.34	0.33	medium
1-5 years	7 (23.3)	4 (26.7)	3 (20.0)			
11-20 years	10 (33.3)	6 (40.0)	4 (26.7)			
21 or more years	6 (20.0)	1 (6.7)	5 (33.3)			
6-10 years	7 (23.3)	4 (26.7)	3 (20.0)			
* Effect size Cohen's d, the others are Cohen's omega. Cohen's omega chosen over phi coefficient due to zero cells						

Table 1. Pharmacist demographics.

Study population

As shown in Table 1, 30 pharmacists were included in the final analysis (black-box AI and uncertainty-aware AI condition). We investigated demographic characteristics for age, gender, race, practice settings, and working years. All the variables were well-balanced according to p-values (*P*>.05). However, according to the effect sizes, some variables were unbalanced. Age had the largest value (0.68). Gender had the only true “small” effect size, of 0.07. The rest were within the range of 0.07 and 0.42.

Outcomes

Participants' overall performance was different across the three conditions regardless of whether the AI and/or the participant was correct (*P*<.001) (Figure 4). The same was true for reaction time.

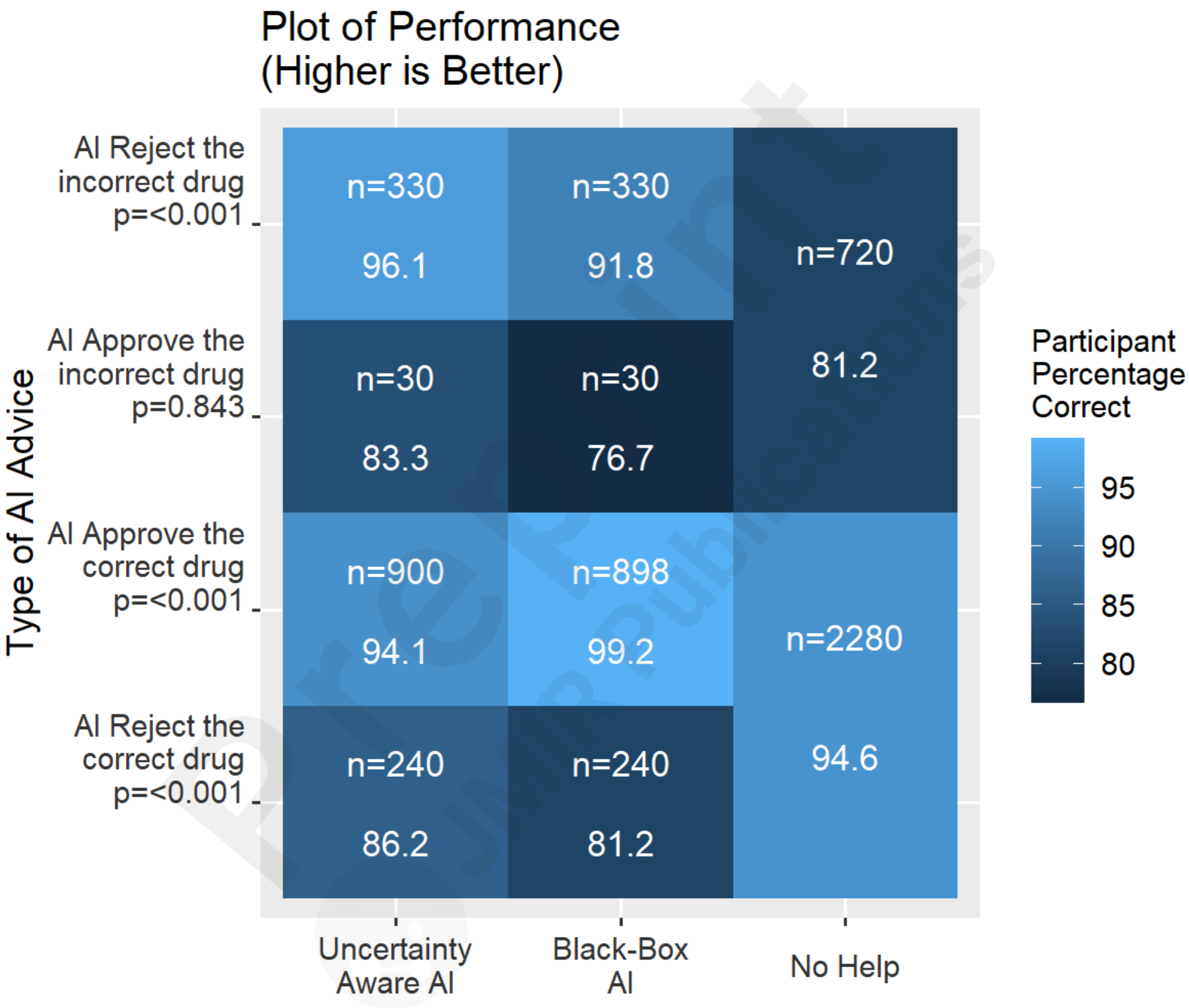


Figure 4. Plot of pharmacists' decision-making performance (% of correct decisions)

AI advice was divided into "Good advice" and "Bad advice". For "Good advice", two subtypes were identified: "AI Rejects Incorrect Drug" (Figure 4, Row 1) and "AI Approves Correct Drug" (Figure 4, Row 3). In the first subtype, pharmacists rejected the incorrect drug 96.1% and 91.8% of the time for uncertainty-aware AI and black-box AI respectively, compared to no AI help (81.2%). When the AI suggested approving the correct drug, performances with black-box AI help surpassed those with uncertainty-aware AI help and

no AI help (99.2% vs 94.1% vs 94.6%).
“AI Approves Incorrect Drug” (Figure 4, Row 2) and “AI Rejects Correct Drug” (Figure 4, Row 4) were classified in the Bad advice category. Despite AI's incorrect predictions, pharmacists demonstrated superior outcomes with uncertainty-aware AI compared to black-box AI help. Uncertainty-aware AI help protected against AI advice suggesting to approve an incorrectly filled medication in contrast with black-box AI (83.3% vs 76.7%). False alerts from AI to reject the correct drug, degraded pharmacists’ performance as compared to baseline performance without AI help. Accuracy for no AI help, uncertainty-aware AI help, and black-box AI help was 94.6%, 86.2%, and 81.2%, respectively.

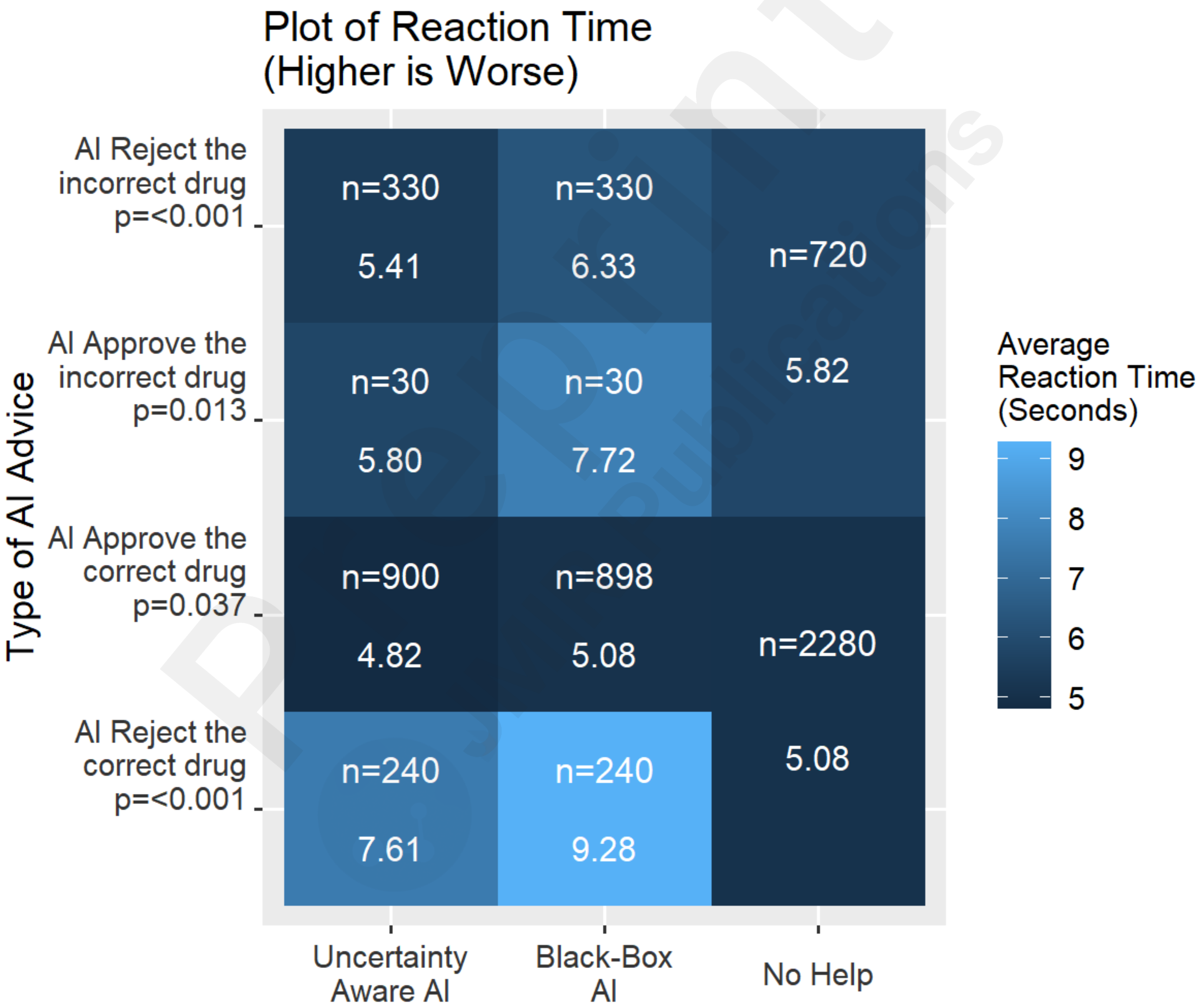


Figure 5 Plot of reaction time (in seconds) of included pharmacists

Figure 5 displays the reaction time for each condition. Pharmacists assisted by uncertainty-aware AI consistently exhibited shorter reaction times than those aided by black-box AI in the four scenarios discussed above (5.41 vs 6.33, 5.80 vs 7.72, 4.82 vs 5.08, 7.61 vs 9.28). Moreover, the uncertainty-aware AI help condition showed improved reaction times compared to the no AI help condition except in the scenario where "AI rejects the correct drug". Black-box AI did not lead to reduced reaction times in any of the situations as contrasted with no AI help.

Discussion

In this study, we performed a randomized control trial to assess the effects of black box and uncertainty-aware AI tools on pharmacist decision-making and reaction time during the medication verification process. In our assessment of the impact of AI assistance on pharmacists' reaction times, distinct patterns emerged between uncertainty-aware AI and black-box AI. Black-box AI slowed down reaction time compared to pharmacists acting alone. In contrast, pharmacists receiving uncertainty-aware AI advice had quicker reaction times when the AI provided accurate predictions, and longer reaction times when the AI produced false positive alerts. Wysocki et al. results mirrored our findings, where AI-user congruence for unexplainable AI resulted in statistically significant differences; AI-user disagreement led to significantly longer reaction times whereas AI-user agreement facilitated quicker decision-making.[28] AI-user agreement also resulted in significantly faster decisions in the explainable AI condition compared to AI advice without an explanation.[28] Our findings and Wysocki et al'.s suggest the transparency and interpretability of uncertainty-aware AI fosters more immediate decision-making when the AI and user agree, and it compels users to closely examine information when they disagree with the AI. These findings underscore the importance of refining AI systems to minimize false alerts to avoid unnecessary double-checks. However, the inability of black-box AI to reduce reaction times across all scenarios prompts concerns about its practical efficacy in fast-paced pharmacy practice.

Arguably, decision accuracy is more important than reaction time due to the patient harm that can result when a bad fill reaches a patient. AI advice, especially uncertainty-aware AI, acted as a protective factor against bad fills, which can improve patient safety. Pharmacists with black-box AI advice, while slower, outperformed uncertainty-aware AI help and no AI help when the AI made correct decisions for correct drug dispensing. Conversely, uncertainty-aware AI advice led to an increase in rejecting correctly filled medications, resulting in unnecessary work for pharmacists and potentially increasing their workload. The uncertainty-aware histogram shown in the uncertainty-aware AI advice was developed based on pharmacists' feedback during the user-centered design phase wherein they expressed a desire to see the probability of each predicted National Drug Code (NDC) to increase the model's transparency.[23] However, these results indicate that the uncertainty-aware AI advice may be confusing or misleading to pharmacists. Further research should explore alternative ways to present AI advice or provide additional user training to help pharmacists correctly interpret uncertainty-aware AI advice.

Obtaining second opinions from colleagues is a common healthcare practice. AI can stand in for colleagues by quickly providing second opinions, but it is critical to communicate any ambiguity or uncertainty in its prediction to foster user trust and allow users to make judgments about when to trust the AI.[10] Previous research has shown that AI errors result in a significant and persistent loss of trust.[29] Explainable AI has been proposed as a solution to increase users' trust by explicitly acknowledging the AI limitations.[10] However, Buçinca et al. found explainable AI may contribute to overreliance on AI advice.[30] Our results indicate that black-box AI can be over-relied upon, especially false negative errors from the AI. A balance between fostering trust and avoiding overreliance is needed to ensure an optimal human-AI teaming experience.

The timing of the AI in relation to the user's decision may influence the user's performance and trust. Gajos and Mamykina recently conducted a study of AI advice timing in 3 conditions: 1) AI recommendation and explanation before making a decision, 2) AI recommendation and explanation after making a decision, and 3) AI explanation only before making a decision.[31] While all 3 conditions led to improved decision accuracy, only the third condition showed improvement in both decision accuracy and learning gain. A 2022 study of AI timing in clinical imaging found veterinary radiologists had lower trust in the AI, lower perceived utility of the AI, and less agreement with the AI, regardless of the AI's correctness, when the advice was presented after their initial clinical decision.[32] Future research should examine the timing of AI advice on pharmacists' trust and accuracy.

Limitations

The study has several limitations. First, the no help condition used a different image set than the AI help conditions. The same image set was used for both AI help conditions. While the images were randomly selected for both AI help and no help conditions, unanticipated biases in the image selection may exist. Second, the images were presented in a fixed order which may have unintentionally biased the results. Future research should randomize the presentation of the images across all conditions for all participants to eliminate any inherent bias. Finally, the small number of participants from two professional pharmacy networks may not be representative of the pharmacy community at large. Future research should be conducted with a broader, more diverse population of pharmacists.

Conclusion

The effectiveness of AI assistance on pharmacists' performance and reaction times varied by AI type and AI accuracy. Overall, uncertainty-aware AI resulted in faster decision-making and black-box AI had the slowest decision-making. Concerningly, black-box AI worsened users' accuracy when the AI provided bad advice, thus increasing the potential for patient harm. Uncertainty-aware AI acted as a safeguard against bad AI advice to approve a misfilled medication, reducing the chance of patient harm. However, pharmacists with uncertainty-aware AI performed worse and had longer reaction times when the AI recommended to incorrectly reject a correctly filled medication compared to no AI help, leading to unnecessary double-checks. These results reinforce the importance of well-designed AI to meet the needs of users to ensure consistent benefits over no AI help.

Acknowledgements

None

References

1. Sahu A, Mishra J, Kushwaha N. Artificial Intelligence (AI) in Drugs and Pharmaceuticals. Comb Chem High Throughput Screen 2022;25(11):1818–1837. PMID:34875986

2. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. Semin Cancer Biol 2021 Jul;72:214–225. PMID:32531273

3. Laves M-H, Ihler S, Ortmaier T, Kahrs LA. Quantifying the uncertainty of deep learning-based computer-aided diagnosis for patient safety. Current Directions in Biomedical Engineering De Gruyter; 2019 Sep 1;5(1):223–226.

4. Khosravi M, Zare Z, Mojtabaeian SM, Izadi R. Artificial Intelligence and Decision-Making in Healthcare: A Thematic Analysis of a Systematic Review of Reviews. Health Serv Res Manag Epidemiol 2024 Mar 5;11:23333928241234863. PMID:38449840

5. Curtain C, Peterson GM. Review of computerized clinical decision support in community pharmacy. J Clin Pharm Ther 2014 Aug;39(4):343–348. PMID:24806361

6. Robertson J, Walkom E, Pearson S-A, Hains I, Williamsone M, Newby D. The impact of pharmacy computerised clinical decision support on prescribing, clinical and patient outcomes: a systematic review of the literature. Int J Pharm Pract 2010 Apr;18(2):69–87. PMID:20441116

7. von Eschenbach WJ. Transparency and the Black Box Problem: Why We Do Not Trust AI. Philos Technol 2021 Dec 1;34(4):1607–1622.

8. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. “hello AI”: Uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. Proc ACM Hum Comput Interact Association for Computing Machinery (ACM); 2019 Nov 7;3(CSCW):1–24.

9. Jussupow E, Spohrer K, Heinzl A, Gawlitza J. Augmenting Medical Diagnosis Decisions? An Investigation into Physicians’ Decision-Making Process with Artificial Intelligence. Information Systems Research INFORMS; 2021 Sep 1;32(3):713–735.

10. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med 2021 Jan 5;4(1):4. PMID:33402680

11. Edmon B, Tanmoy B, Dimitri K. The need for uncertainty quantification in machine-assisted medical decision making. Nature Machine Intelligence; Basingstoke Basingstoke, United States, Basingstoke: Nature Publishing Group; 2019 Jan 20;1(1):20–23.

12. Tomsett R, Preece A, Braines D, Cerutti F, Chakraborty S, Srivastava M, Pearson G, Kaplan L. Rapid Trust Calibration through Interpretable and Uncertainty-Aware AI.

Patterns (N Y) 2020 Jul 10;1(4):100049. PMID:33205113

13. Prabhudesai S, Yang L, Asthana S, Huan X, Liao QV, Banovic N. Understanding Uncertainty: How Lay Decision-makers Perceive and Interpret Uncertainty in Human-AI Decision Making. Proceedings of the 28th International Conference on Intelligent User Interfaces New York, NY, USA: Association for Computing Machinery; 2023. p. 379–396.

14. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012 Jan-Feb;19(1):121–127. PMID:21685142

15. Campbell PJ, Patel M, Martin JR, Hincapie AL, Axon DR, Warholak TL, Slack M. Systematic review and meta-analysis of community pharmacy error rates in the USA: 1993–2015. BMJ Open Qual British Medical Journal Publishing Group; 2018 Oct 1;7(4):e000193.

16. Parasuraman R, Mouloua M. Interaction of signal discriminability and task type in vigilance decrement. Percept Psychophys 1987 Jan;41(1):17–22. PMID:3822739

17. Chui MA, Look KA, Mott DA. The association of subjective workload dimensions on quality of care and pharmacist quality of work life. Res Social Adm Pharm 2014 Mar-Apr;10(2):328–340. PMID:23791360

18. Warm J, Dember WN, Hancock P. Vigilance and workload in automated systems. Automation and human 1996; doi: 10.1201/9781315137957-9/vigilance-workload-automated-systems-joel-warm-william-dember-peter-hancock

19. Manly T, Robertson IH, Galloway M, Hawkins K. The absent mind: further investigations of sustained attention to response. Neuropsychologia 1999 Jun;37(6):661–670. PMID:10390027

20. NCPA Releases 2023 Digest Report. NCPA Releases 2023 Digest Report. Available from: https://ncpa.org/newsroom/news-releases/2023/10/15/ncpa-releases-2023-digest-report [accessed May 16, 2024]

21. Karia A, Norman R, Robinson S, Lehnbom E, Laba T-L, Durakovic I, Balane C, Joshi R, Webster R. Pharmacist's time spent: Space for Pharmacy-based Interventions and Consultation TimE (SPICE)—an observational time and motion study. BMJ Open British Medical Journal Publishing Group; 2022 Mar 1;12(3):e055597. PMID:35236731

22. Hohmeier KC, Wheeler J, Heintz K, Gatwood J. Community pharmacist workflow and medication therapy management delegation: An assessment of preferences and barriers. J Am Pharm Assoc 2020 Aug 16;60(6):e215–e223. PMID:32811750

23. Zheng Y, Rowell BE, Chen Q, Kim JY, Kontar R, Yang XJ, Lester CA. Designing human-centered AI to prevent medication dispensing errors: Focus group study with pharmacists. JMIR Form Res 2023 Aug 17;7:e51921. PMID:38145475

24. Lester CA, Li J, Ding Y, Rowell B, Yang J 'xi', Kontar RA. Performance evaluation of a prescription medication image classification model: an observational cohort. NPJ Digit Med 2021 Jul 27;4(1):118. PMID:34315995

25. Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Balcan MF, Weinberger KQ, editors. New York, New York, USA: PMLR; 20--22 Jun 2016;48:1050–1059.

26. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016. p. 770–778.

27. Team RC. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http s. www R-proje ct org 2017;

28. Wysocki O, Davies JK, Vigo M, Armstrong AC, Landers D, Lee R, Freitas A. Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI-driven clinical decision-making. Artif Intell 2023 Mar 1;316:103839.

29. Chong L, Zhang G, Goucher-Lambert K, Kotovsky K, Cagan J. Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. Comput Human Behav 2022 Feb 1;127:107018.

30. Buçinca Z, Malaya MB, Gajos KZ. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. Proc ACM Hum-Comput Interact New York, NY, USA: Association for Computing Machinery; 2021 Apr 22;5(CSCW1):1–21.

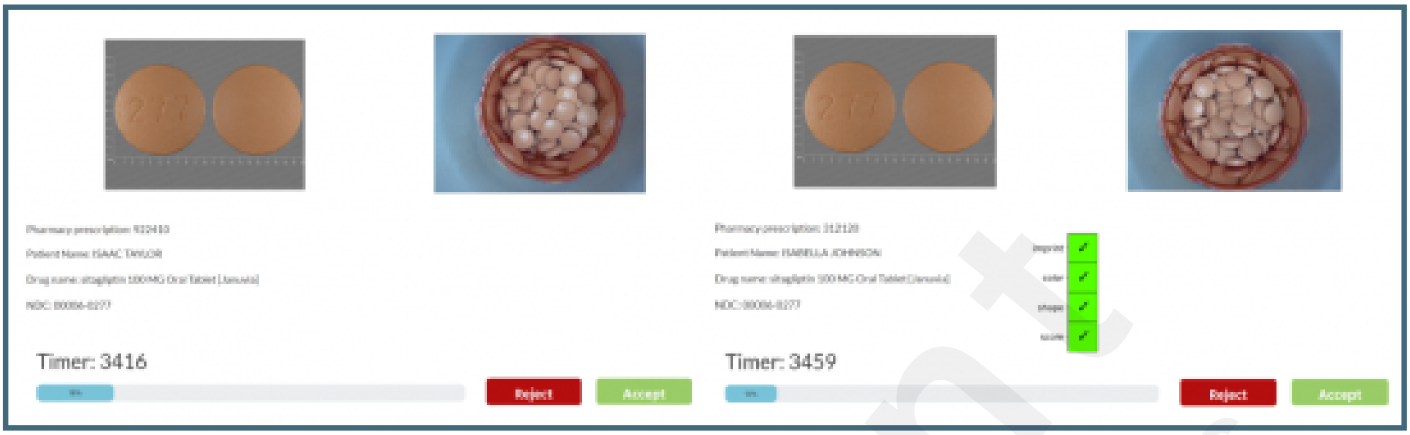
31. Gajos KZ, Mamykina L. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. 27th International Conference on Intelligent User Interfaces New York, NY, USA: Association for Computing Machinery; 2022. p. 794–806.

32. Fogliato R, Chappidi S, Lungren M, Fisher P, Wilson D, Fitzke M, Parkinson M, Horvitz E, Inkpen K, Nushi B. Who Goes First? Influences of Human-AI Workflow on Decision Making in Clinical Imaging. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency New York, NY, USA: Association for Computing Machinery; 2022. p. 1362–1374.

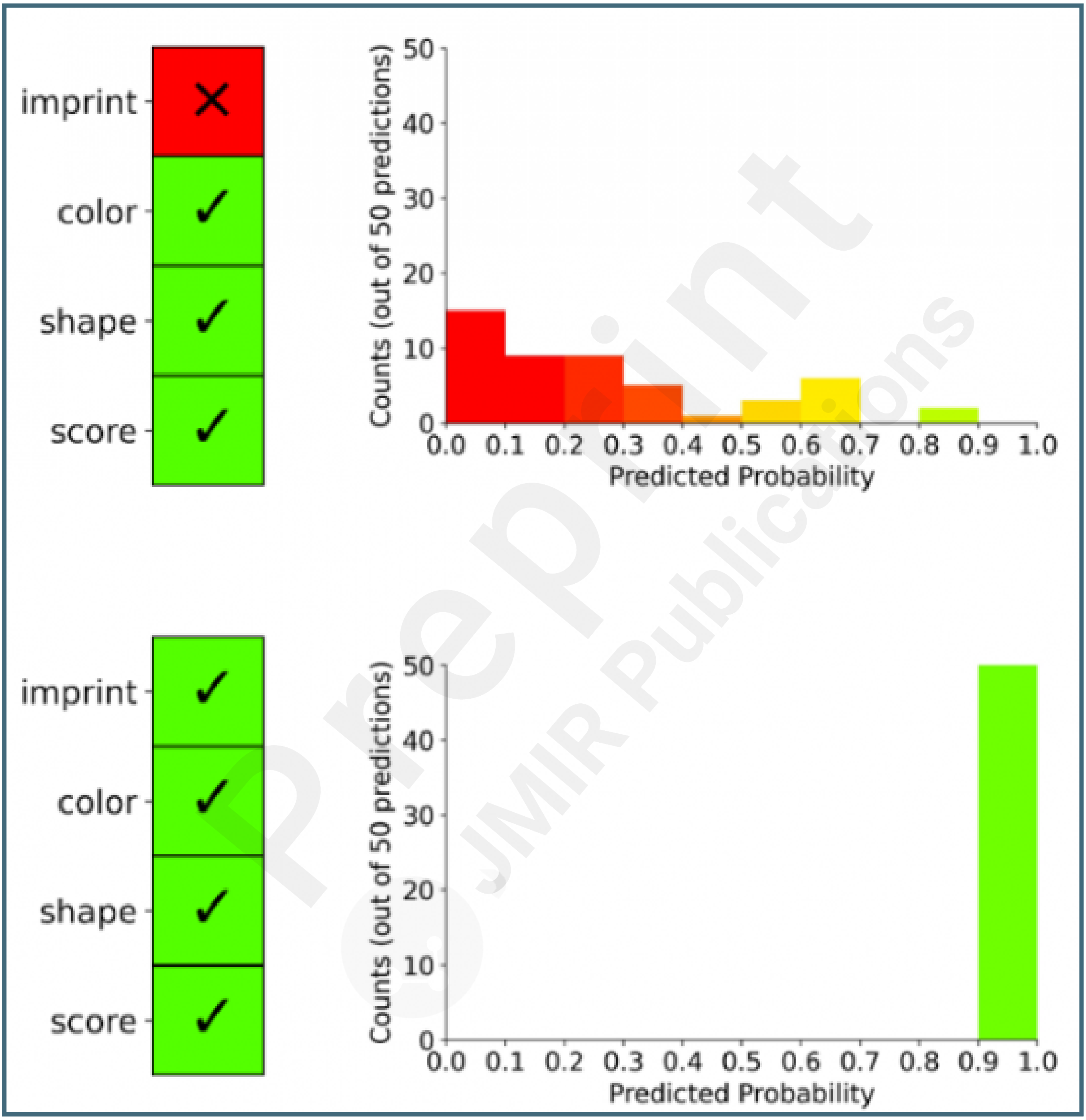
Supplementary Files

Figures

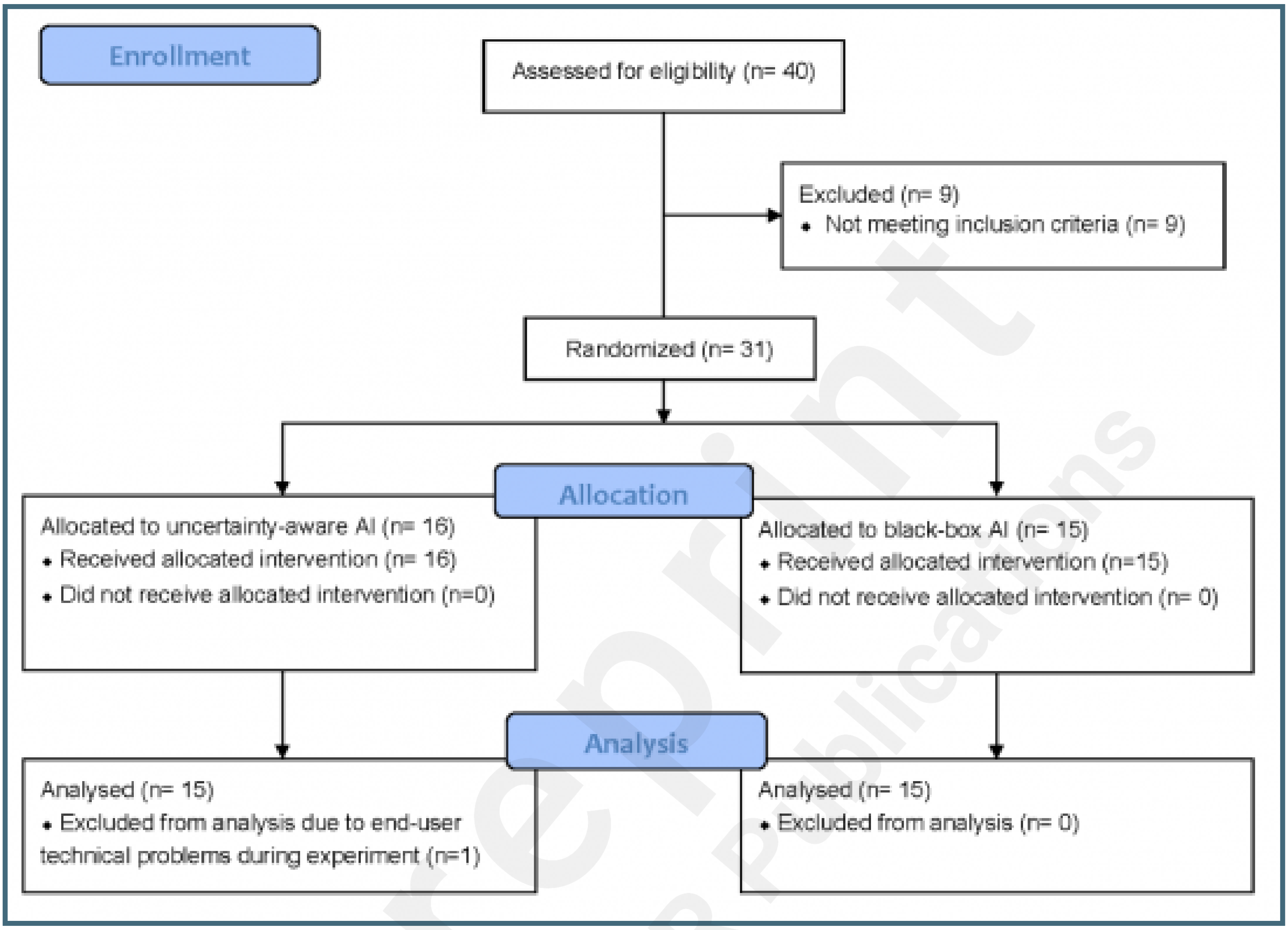
An example of the interface. The task was to compare the “reference image” on the left to the “fill image” on the right to determine whether to accept or reject the dispensed medication. The “reference image” is the ground truth. For half the trials there was AI help on the bottom right of the screen.



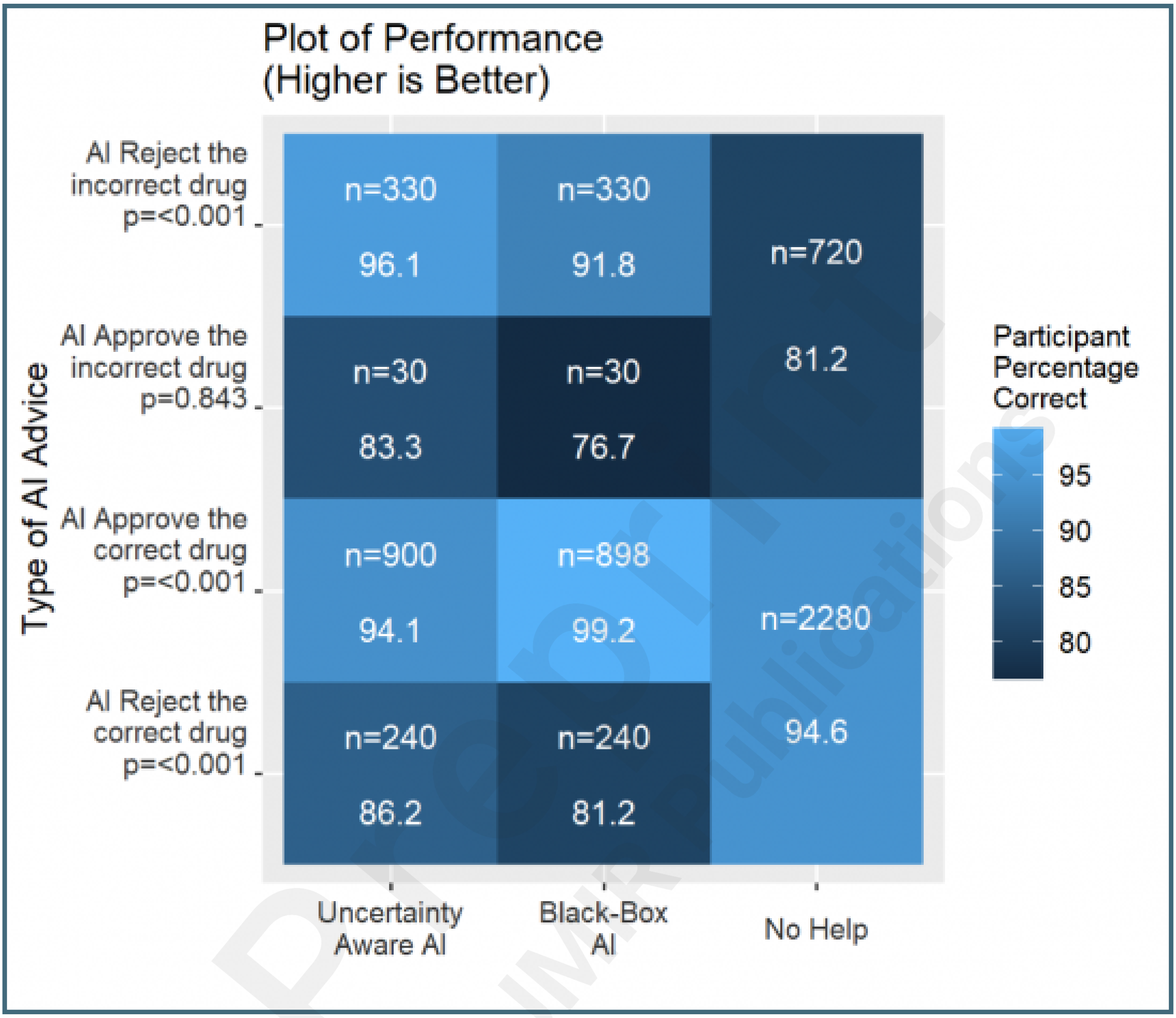
Presentations of two AI models. The left-hand graphs are “black-box AI” which displays “yes” or “no” for four pill characteristics: imprint, color, shape, and score. The right-hand graphs are “uncertainty-aware AI” which illustrates the predicted probability of each set of predictions summarized in a histogram.



Participant flow of the RCT for effect of artificial intelligence models on medication dispensing in pharmacists.



Plot of pharmacists' decision-making performance (% of correct decisions).



Plot of reaction time (in seconds) of included pharmacists.