Preventing medication dispensing errors in pharmacy practice with interpretable machine intelligence – Images Wave 3

Protocol Number: HUM00241223

NCT: NCT06245044

Principal Investigator: Corey A. Lester, PharmD, PhD

Funded by: National Library of Medicine

Version Number: v.2.0

26 August 2025

Table of Contents

1 PROTOCOL SUMMARY

1.1 SYNOPSIS

Title:	Preventing medication dispensing errors in pharmacy practice with interpretable machine intelligence – Images Wave 3	
Study Description:	This study is the third wave of the participatory design process and a formative evaluation to refine the human-machine intelligence (MI) collaboration for the pharmacy verification task in a simulated environment.	
Objectives:	Primary Objectives:	Determine the impact of MI assistance on task time, error detection rate, and users' trust
	Secondary Objectives:	Examine the effect of MI assistance on cognitive effort
Endpoints:	Primary Endpoints:	Task time, decision accuracy, and trust
	Secondary Endpoints:	Cognitive effort
Study Population:	Licensed pharmacists in the United States with medication dispensing experience who are 18 or older.	
Study Duration:	12 months	
Participant Duration:	1 day	

2 INTRODUCTION

2.1 STUDY RATIONALE

Medication errors are a major public health problem. Preventable medication errors cost the U.S healthcare system over $16 billion annually with $4.2 billion attributable to preventable outpatient medication errors.[1] Medication errors result in 3 million outpatient medical appointments, 1 million emergency department visits, and 125,000 hospital admissions each year.[2] In addition to the added costs and strain on the healthcare system, medication errors can result in liability claims against pharmacy staff. The two most common reasons for legal action against licensed pharmacy staff is dispensing of the wrong dose or wrong medication which results in an average settlement of nearly $125,000.[3] There is a critical need to provide pharmacy staff with tools to reduce medication errors thereby improving patient safety and reducing costs.

Pharmacists currently perform an independent double-check to identify drug-selection errors before they can reach the patient. However, the use of machine intelligence (MI) to support this cognitive decision-making work by pharmacists does not exist in practice. Instead, pharmacists rely solely on reference images of the medication which they can compare to the prescription vial contents. Previous research has shown that decision support systems can effectively improve healthcare delivery efficiency and accuracy, while preventing adverse drug events.[4] However, little is known about how MI technologies impact pharmacists' work performance and cognitive demand.

To facilitate the long-term symbiotic relationship between the pharmacists and the MI system, proper trust needs to be established. While trust has been identified as the central factor for effective human-machine teaming,[5–15] issues arise when humans place unjustified trust in automated technologies do not place enough trust in them.[7,8,10,12–15] Over trust in automation can lead to complacency[9,16] and automation bias.[17,18] For instance, the pharmacists may rely on the MI system to the extent that they blindly accept any recommendation by the system. Under trust can result in pharmacist disuse and potential abandonment of the MI system.[19]

Furthermore, little is known about the timing of the MI advice on pharmacists' work performance. For example, showing the MI's advice while the pharmacist is performing the medication verification task may yield different results than showing the MI's advice after the pharmacist made their decision.

The study investigators have developed a MI system for medication images classification. The objective of this study is to examine the effectiveness of the timing of MI advice to determine if it results in lower task time, increased accuracy, and increased trust in the MI. The objectives of this study are to: 1) evaluate changes in participants' trust due to the use of interpretable machine intelligence; and 2) determine the effect of interpretable machine intelligence on participants' performance in a simulated environment.

3 OBJECTIVES AND ENDPOINTS

OBJECTIVES	ENDPOINTS
Primary	
Determine the impact of MI assistance on task time, decision accuracy rate, trust change and trust.	• Difference in task time measured by the number of seconds from starting the task to accepting or rejecting a medication image • Difference in detection rate measured by number of medication verification errors • Difference in trust measured by visual analog trust scale and the trust scale
Secondary	
Determine the effect of MI assistance on cognitive effort, work load, and usability	• Difference in cognitive effort measured by duration of fixation • Difference in cognitive effort as measured by fixation count • Difference in cognitive effort as measured by dwell time. Dwell time is calculated by summing up the duration of fixations in a region (e.g., the fill image) for each trial. • NASA task load • SUS

4 STUDY DESIGN

4.1 OVERALL DESIGN

This study will examine the effectiveness of MI assistance on the medication verification process in a mock setting. Pharmacists with medication dispensing experience will be recruited to perform mock verification tasks on a web application on the computer. Webcam-based eye-tracking software will be used to detect eye gaze patterns during verification tasks (*https://www.labvanced.com/*).

Participants will attend one 3-hour remote visit. During this visit, participants will complete a series of 300 remote verification tasks. Prior to starting the verification tasks, participants will complete 3 surveys: 1) a demographics survey 2) the propensity to trust scale[20] and 3) the Occupational Fatigue Exhaustion Recovery Scale.[21] Each participant will then receive a 15-minute self-guided training session that introduces the study and teaches participants how to use the interface and perform the medication verification task. In each verification task, participants will compare the image on the left (i.e., the medication filled for a prescription) to the image on the right (i.e., a known reference image) and decide whether to accept or reject the images as containing the same prescribed medication.

Pharmacists will be presented with 3 conditions: No help group, Scenario #1, and Scenario #2 (see Figure 1). The order of the conditions will be randomized. In the No help group, participants will complete the verification task without any artificial intelligence (AI) help. In Scenario #1, participants will receive AI in the form of a pop-up message if their decision differs from the AI's determination. In Scenario #2, AI help will be displayed concurrently with the filled and reference images.

After each verification task in Scenario #1 and Scenario #2, the participants will be asked to respond to a visual analog trust scale (e.g., How much do you trust the automation with the left anchor "I completely distrust the MI" and the right anchor "I completely trust the MI"?).[20,22] After each condition, participants will complete 2 brief surveys - System Usability Scale,[23] NASA Task Load Index.[24] After Scenario #1 and Scenario #2, participants will also complete a trust scale.[25]

To approximate a real-world experience in this mock setting, 20% of images provided in these verification tasks for each participant will differ from labels in prescriptions (i.e., an incorrectly filled prescription).

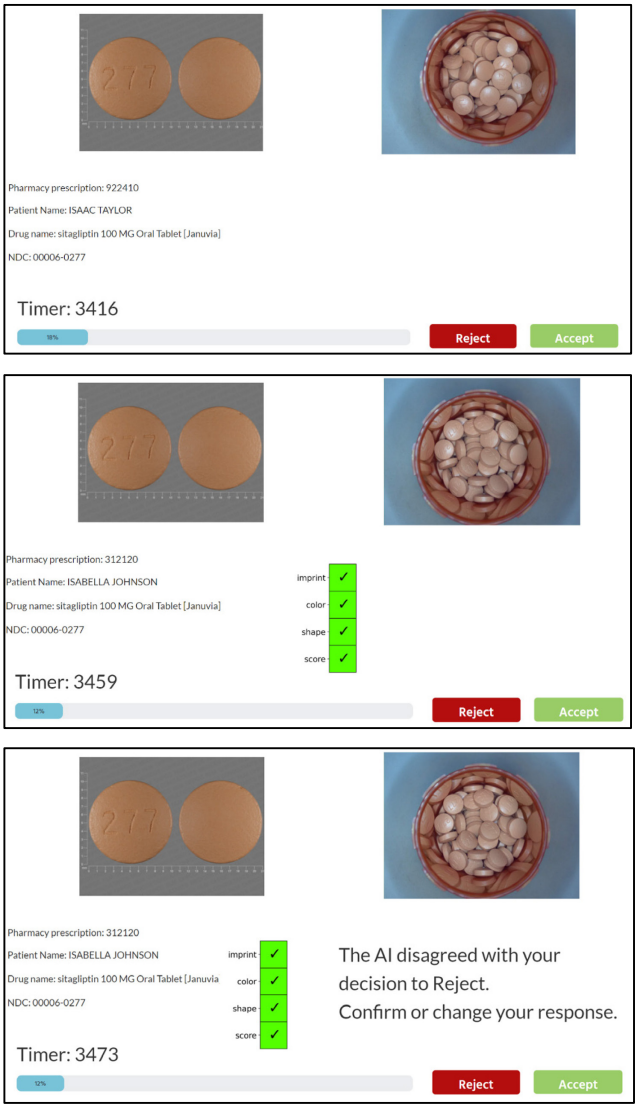

Figure 1. Interfaces for each study condition evaluating the effect of machine intelligence (MI) on work performance and pharmacist trust. Top: No MI help; Middle: MI help; Bottom: AI help when disagreement between user and AI

5 STUDY POPULATION

5.1 INCLUSION CRITERIA

To be eligible to participate in this study, an individual must meet all the following criteria:

1. Licensed pharmacist in the United States
2. Age 18 years and older at screening
3. PC/Laptop with Microsoft Windows 10 or Mac (Macbook, iMac) with MacOS with Google Chrome, Edge, Opera, Safari, or Firefox web browser installed on the device

4. Screen resolution of 1024x968 pixels or more
5. A laptop integrated webcam or USB webcam is also required for the eye tracking purpose.

5.2 EXCLUSION CRITERIA

1. Participated in Wave 1 or Wave 2
2. Eyeglasses
3. Uncorrected cataracts, intraocular implants, glaucoma, or permanently dilated pupil
4. Require a screen reader/magnifier or other assistive technology to use the computer
5. Eye movement or alignment abnormalities (lazy eye, strabismus, nystagmus)

5.3 STRATEGIES FOR RECRUITMENT AND RETENTION

Pharmacists will be recruited from 3 recruitment pools of pharmacists who have expressed interest in this study or in research in general. They include the Minnesota Pharmacy Practice-Based Research Network, PearlRx (a University of Wisconsin-Madison based pharmacy practice-based research network), and the University of Michigan College of Pharmacy Pharmacist Preceptor Network. These 3 recruitment pools have more than 3000 registered pharmacists who may be eligible to participate in the study.

Email communication will be delivered to advertise the study and help ensure adequate participation in the study to reach recruitment goals. Participants will receive $100 for completing the study visit.

Participation is limited to one remote study visit. As such, we do not anticipate any need for a formal retention plan.

6 STUDY INTERVENTION

6.1 MEASURES TO MINIMIZE BIAS: RANDOMIZATION AND BLINDING

The order of conditions will be randomization will occur using Labvanced's customizable randomization scheme. This will be a stratified random sample with balanced assignment to each study condition.

This is an unblinded study. Participants and researchers will know the randomization group.

7 STATISTICAL CONSIDERATIONS

7.1 STATISTICAL HYPOTHESES

- Primary Endpoint(s):

We hypothesize reaction time and accuracy will be different between the three scenarios.

We hypothesize trust will be a difference between scenario 1 and scenario 2.

- Secondary Endpoint(s):

We hypothesize there will be a difference in cognitive effort, workload, and usability between the three scenarios.

7.2 SAMPLE SIZE DETERMINATION

For the within-subject design to obtain 95% power with a 95% confidence interval, 36 subjects are required to detect a statistically significant difference in detection rates between the experimental groups.

7.3 STATISTICAL ANALYSES

7.3.1 ANALYSIS OF THE PRIMARY AND SECONDARY ENDPOINT(S)

ANOVA and/or mixed linear and mixed generalized linear models will be used to test the difference in task time, trust, detection rate, and cognitive effort among the three study conditions. The family-wise error rate will be set at a value of no greater than 0.05 to account for the 3 pairwise comparisons. ANOVA and Fisher's exact test are used when comparing independent proportions. All statistics will be performed using R (version 3.6.1). The ANOVA test and Fisher's exact test will be computed with the stats library.

8 REFERENCES

1. Hong K, Hong YD, Cooke CE. Medication errors in community pharmacies: The need for commitment, transparency, and research. *Res Social Adm Pharm*. 2019;15(7):823-826.

2. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. *National Action Plan for Adverse Drug Event Prevention*.; 2014.

3. Reiner G, Pierce SL, Flynn J. Wrong drug and wrong dose dispensing errors identified in pharmacist professional liability claims. *J Am Pharm Assoc* . Published online March 23, 2020. doi:10.1016/j.japh.2020.02.027

4. Landerdahl M. *Overview of the Utilisation of Automated Decision Support Tools at Dispensing in the Pharmacies in the European Countries*. Masters. University of Gothenburg; 2010. https://www.pcne.org/upload/files/35_MedRev_Landerdahl.pdf

5. Lu Y, Sarter N. Eye tracking: A process-oriented method for inferring trust in automation as a function of priming and system reliability. *IEEE Trans Hum Mach Syst*. 2019;49(6):560-568.

6. Lee JD, See KA. Trust in automation: Designing for appropriate reliance. *Hum Factors*. 2004;46(1):50-80.

7. Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human-robot interaction. *Hum Factors*. 2011;53(5):517-527.

8. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. *Hum Factors*. 2015;57(3):407-434.

9. Wickens CD, Clegg BA, Vieane AZ, Sebok AL. Complacency and automation bias in the use of imperfect automation. *Hum Factors*. 2015;57(5):728-739.

10. Robinette P, Li W, Allen R, Howard AM, Wagner AR. Overtrust of robots in emergency evacuation scenarios. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE; 2016. doi:10.1109/hri.2016.7451740

11. Lee JJ, Knox WB, Wormwood JB, Breazeal C, Desteno D. Computationally modeling interpersonal trust. *Front Psychol*. 2013;4:893.

12. Joslyn SL, LeClerc JE. Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J Exp Psychol Appl*. 2012;18(1):126-140.

13. Booth S, Tompkin J, Pfister H, Waldo J, Gajos K, Nagpal R. Piggybacking robots. In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM; 2017. doi:10.1145/2909824.3020211

14. Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. *Hum Factors*. 1997;39(2):230-253.

15. Hoffman RR, Johnson M, Bradshaw JM, Underbrink A. Trust in Automation. *IEEE Intell Syst*. 2013;28(1):84-88.

16. Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theor Issues Ergon*. 2007;8(4):321-348.

17. Mosier KL, Skitka LJ, Heers S, Burdick M. Automation bias: decision making and performance in high-tech cockpits. *Int J Aviat Psychol*. 1997;8(1):47-63.

18. Cummings M. Automation bias in intelligent time critical decision support systems. In: *AIAA 1st Intelligent Systems Technical Conference*. American Institute of Aeronautics and Astronautics; 2004. doi:10.2514/6.2004-6313

19. Bliss J, Dunn M, Fuller BS. Reversal of the cry-wolf effect: An investigation of two methods to increase alarm response rates. *Percept Mot Skills*. 1995;80(3_suppl):1231-1242.

20. Merritt SM, Heimbaugh H, LaChapell J, Lee D. I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. *Hum Factors*. 2013;55(3):520-534.

21. Winwood PC, Winefield AH, Dawson D, Lushington K. Development and validation of a scale to measure work-related fatigue and recovery: the Occupational Fatigue Exhaustion/Recovery Scale (OFER). *J Occup Environ Med*. 2005;47(6):594-606.

22. Wiczorek R, Manzey D. Supporting attention allocation in multitask environments: effects of likelihood alarm systems on trust, behavior, and performance. *Hum Factors*. 2014;56(7):1209-1221.

23. Brooke J. SUS: A quick and dirty usability scale. *Usability evaluation in industry*. 1995;189(194). Accessed August 23, 2023. https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale

24. NASA Task Load Index. Accessed August 23, 2023. https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/nasa-task-load-index

25. Muir BM, Moray N. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*. 1996;39(3):429-460.