Using Discrete Event Simulation to Design and Assess an AI-aided Workflow for Same-day Diagnostic Testing of Women Undergoing Breast Screening

Yannan Lin, MD, MPH, PhD[1], Anne C. Hoyt, MD[2], Vladimir G. Manuel, MD[3,4], Moira Inkelas, MPH, PhD[4,5], William Hsu, PhD[1,6]

[1]Medical & Imaging Informatics, Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; [2]Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; [3]Department of Family Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; [4]UCLA Clinical and Translational Science Institute, Los Angeles, CA, USA; [5]Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, CA, USA; [6]Department of Bioengineering, University of California, Los Angeles, CA, USA

Abstract

The process of patients waiting for diagnostic examinations after an abnormal screening mammogram is inefficient and anxiety-inducing. Artificial intelligence (AI)-aided interpretation of screening mammography could reduce the number of recalls after screening. We proposed a same-day diagnostic workup to alleviate patient anxiety by employing an AI-aided interpretation to reduce unnecessary diagnostic testing after an abnormal screening mammogram. However, the potential unintended consequences of introducing this workflow in a high-volume breast imaging center are unknown. Using discrete event simulation, we observed that implementing the AI-aided screening mammogram interpretation and same-day diagnostic workflow would reduce daily patient volume by 4%, increase the time a patient would be at the clinic by 24%, and increase waiting times by 13-31%. We discuss how changing the hours of operation and introducing new imaging equipment and personnel may alleviate these negative impacts.

Introduction

Breast cancer is the most common cancer and the second leading cause of cancer-related death among women in the United States. (1) In randomized clinical trials, screening mammography has reduced breast cancer-related mortality. (2-4) The American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) is designed to facilitate appropriate description, categorization, and assignment of management recommendations of mammographic findings. When an incomplete assessment (category 0) is assigned at screening, additional imaging evaluation and/or comparison with prior examination(s) is recommended before a final assessment category can be rendered. (5) Our institution's recall rate ranges from 7-16% (mean 11%), varying across breast radiologists depending on their years of experience. Most recalled women, after undergoing diagnostic mammography and/or diagnostic ultrasound, are found to have benign findings that do not require further testing. After ten annual screening mammograms, there is an up to 60% chance of having at least one false positive result. (6, 7). Simulation studies have demonstrated that AI-aided interpretation reduces the number of recommended recalls after screening mammography. (8)

At most breast imaging centers, women are discharged home after their screening mammogram. Screening mammograms are read "offline" in batches at a later date rather than in a real-time "online" manner. Patients who are assigned a BI-RADS category 0 are notified and scheduled for a diagnostic imaging examination at a later date. This waiting period between the screening and diagnostic examination is associated with increased anxiety, and studies have shown that waiting time triggers patient anxiety. (9) (10) Providing immediate confirmation and same-day further evaluation of a suspicious finding would lower anxiety in patients undergoing screening mammography. (11) Besides anxiety, longer waiting time for a diagnostic evaluation may delay cancer diagnosis, especially if the patient does not return for the recommended additional evaluation. Furthermore, some patients may be unable to travel to a breast imaging center for further evaluation, identifying a particularly vulnerable subset of women. Additional communication may be needed to ensure high adherence with a timely diagnostic follow-up examination (12), introducing additional workload to the staff. Some clinics, such as Baylor Health, Robert Wood Johnson-Barnabas, and Mass General Brigham, have implemented same-day diagnostic testing at a single site and have reported increased patient satisfaction. (13, 14) Recent advances in artificial intelligence (AI) have yielded algorithms that can interpret screening mammograms as well as or better than radiologists. (15)

To address these potential harms of breast screening, we propose a revised workflow for patients undergoing screening mammography. First, we plan to implement an immediate AI-aided interpretation of the screening mammograms in all patients. If AI determines the result is likely normal (i.e., BI-RADS 1 or 2), the woman is notified that the examination appears normal and is allowed to leave the center. Our current workflow batches these "AI normal" examinations for later interpretation (e.g., the next day). However, if the AI algorithm determines that the screening mammogram is abnormal (i.e., BI-RADS 0), a radiologist will immediately confirm the need for further diagnostic workup, and if verified, the patient will undergo same-day diagnostic imaging examinations. The scheduling of examinations at an imaging center needs to accommodate these additional same-day diagnostic tests and would be facilitated by an onsite patient coordinator. Thus, evaluating the impact of the AI algorithm to mitigate any adverse effects (e.g., longer waiting times) in terms of providing same-day diagnostic examinations is imperative before on-site implementation.

A simulation represents the functioning of a real-world system as it operates over time. It can identify potential bottlenecks and answer "what-if" questions about real-world situations without practical and/or financial ramifications. (16-19) Discrete event simulation (DES) is a technique that models the operation of a system as a discrete series of events in time, aiming to evaluate, predict, and optimize an existing or proposed system. (16) DES has been widely used in healthcare settings to support better operational decision-making and planning. (20, 21) In this study, we created a baseline DES model reflecting the current patient flow at one of our largest breast imaging centers. Next, we designed an AI-aided workflow model incorporating the same-day diagnostic breast imaging workup after patients undergo breast screening mammograms. We then assessed the potential impacts of implementing the new workflow on patient volume, length of stay, and waiting times by answering what-if questions. The same-day diagnostic workup affects the radiologists' workflow; our study focuses on the patient experience. The findings of this study provide the basis for future studies analyzing patient satisfaction, recall rate reduction, and efficiency of breast screening programs.

Methods

Institutional Review Board (IRB) approval was obtained at the University of California, Los Angeles (IRB#23-001126) to collect study data as part of health care. The development and validation of the DES models in this study followed a four-step procedure, including planning, modeling, verification/validation, and analysis. (22) We detail each step in the following sections.

1. Planning

The Barbara Kort Women's Imaging Center (BKWIC) is one of the largest breast imaging centers at UCLA Health, performing ~22% of all breast imaging examinations and procedures across all sites. The patient may come in for one of six examinations or procedures, including screening mammography, both diagnostic mammography and diagnostic ultrasound, diagnostic mammography only, diagnostic ultrasound only, stereotactic (mammogram)-guided core needle biopsy, or ultrasound-guided core needle biopsy. The baseline workflow model is designed to model the patient flow in this imaging center. The proposed workflow includes additional steps, including AI-aided interpretation of screening mammograms and same-day diagnostic imaging. In the planning stage, we engaged stakeholders with expertise in breast imaging (A.H.), AI (W.H., Y.L.), and implementation science (V.M. and M.I.) to define the system and understand the workflow in the breast imaging center. Two authors (Y.L. and V.M.) completed half-day visits to the imaging center under the supervision of a breast radiologist (A.H.) to shadow staff, technologists, and breast radiologists, with the goals of understanding their work responsibilities and timing each step that contributed to the estimation of DES model parameters. While shadowing, efforts were made to minimize disruptions to imaging center personnel's daily tasks and avoid direct contact with the patients.

2. Modeling

The structure of the baseline model reflects the patient flow in the center (see **Figure 1-A).** Resources shared among all patients include check-in staff, public waiting room, dressing room, and gowned waiting room. The consent room and relevant personnel are shared among biopsy patients. Mammography equipment and technologists are shared across patients who come in for a screening mammogram, diagnostic mammogram, or stereotactic-guided biopsy. In contrast, ultrasound units are shared across patients undergoing diagnostic ultrasound or ultrasound-guided biopsy. Each imaging or biopsy step contains multiple components described in **Table S1** in the Supplement. In the revised workflow (see **Figure 1-B**), the screening mammography step is followed by an AI assessment step where an AI algorithm interprets the screening mammogram and outputs a malignancy risk score between 1 and 10. Patients with AI-risk scores between 8 and 10 and confirmed to be suspicious by the radiologist are deemed eligible for same-day diagnostic imaging. They will undergo one of the three diagnostic workups, including (1) diagnostic mammography and diagnostic ultrasound, (2) diagnostic mammography only, or (3) diagnostic ultrasound only. We assumed patient arrivals followed the Poisson process, the length of stay in each step followed a normal distribution, and the

proportions of different types of patients each day were time-invariant. We assumed the radiologists were not bottlenecks in the system by merging their time with technologists reviewing diagnostic images into the imaging or procedure step. In the proposed workflow, the radiologists maintained regular worklists for screening, diagnostics, and biopsies, assuming the same-day workup is integrated into their existing workflow. Model parameters were estimated using three-day (August 18, 21, and 22, 2023) scheduling data, one-year productivity data (June 2022 to May 2023), and time-motion data obtained during clinic shadowing. **Table 1** describes each step and the corresponding parameters.
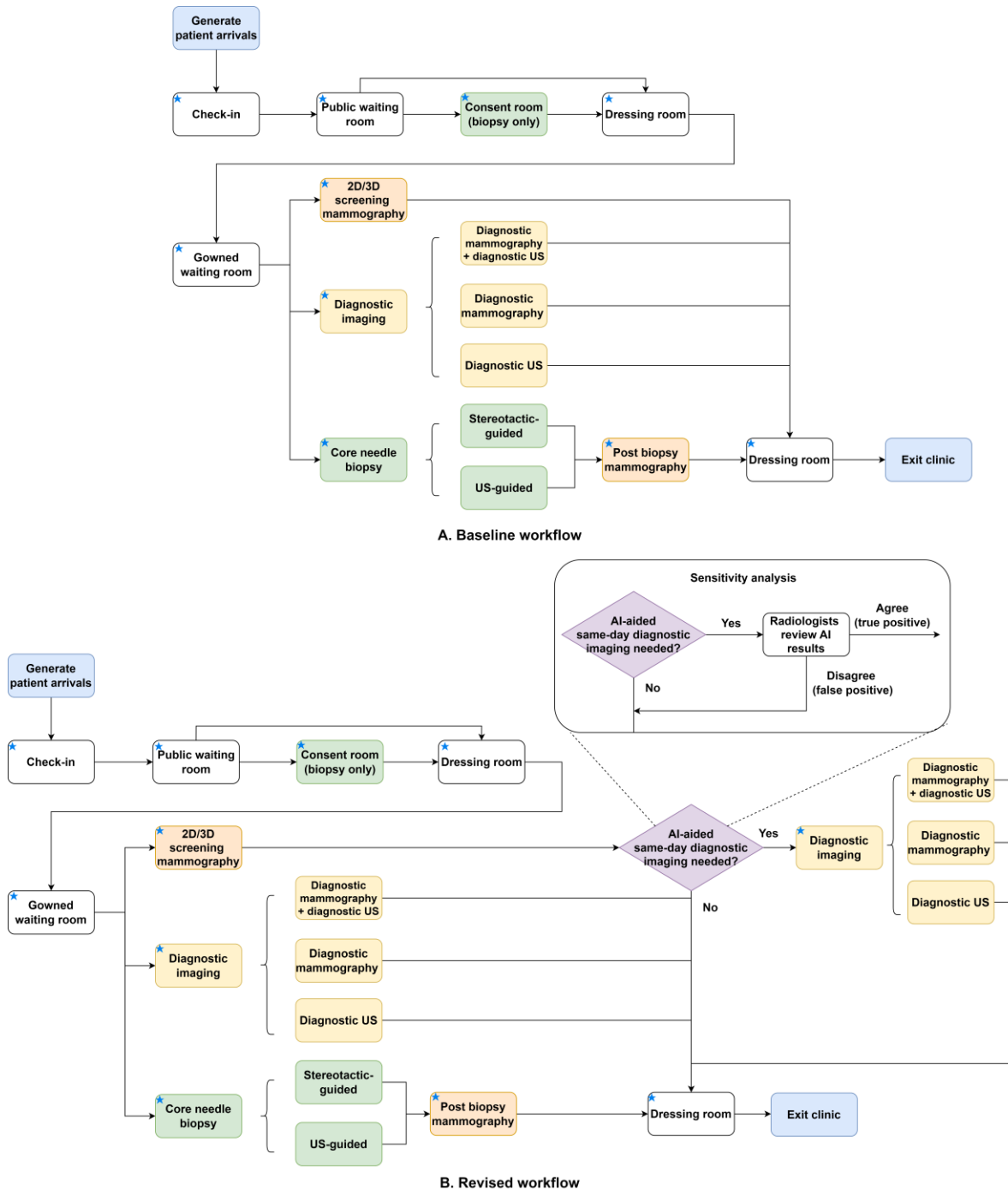

Figure 1. The baseline and revised workflows in a breast imaging center. **Notes**: A blue star in the box's upper left corner indicates that device and/or personnel resources are associated with this step. **Abbreviations**: 2D: two-dimensional; 3D: three-dimensional; AI: artificial intelligence; US: ultrasound.

Table 1. Parameters of discrete event simulation models.

Name	Personnel	Capacity	Condition	Distribution of length of stay (hour)
Operating hours	NA	NA	Stops patient check-in at 8.5 hours and closes at 9 hours	NA
Patient arrival	NA	NA	Arrival rate: 6 per hour Mean interarrival time: 1/arrival rate	Poisson process
Patient check-in	Staff	3	NA	Normal (0.05, 0.01)
Public waiting room	NA	20		Normal (0.17, 0.034)
Consent room	LVN/trainee	1	NA	Normal (0.17, 0.034)
Dressing room	NA	3	NA	Normal (0.03, 0.006)
Gowned waiting room	NA	5	NA	Normal (0.017, 0.0034)
AI assessment (revised workflow only)	Radiologist	NA	The percentage of the screener needs a diagnostic imaging examination Normal (12%, 5%) Sensitivity analysis: AI recall rate: uniform (0.1, 0.2, 0.3) Radiologists agree with the AI rate: uniform (0.7, 0.8, 0.9)	Normal (0.25, 0.05) Sensitivity analysis: AI: normal (0.05, 0.01) Radiologist: normal (0.20, 0.04)
2D/3D Screening mammography	Mammo technologist	3 (mammo);	50.3% screeners/day	Normal (0.17, 0.034)
Diagnostic imaging	Mammo/US technologist	2 (US)	45.0% diagnostic imaging patients/day (Diagnostic mammo + diagnostic US: 80%; Diagnostic US: 10%; Diagnostic mammo: 10%)	Diagnostic mammo: normal (0.5, 0.1); Diagnostic US: normal (0.5, 0.1)
Core needle biopsy	Mammo/US technologist, radiologist, LVN/trainee		4.0% US-guided biopsy/day 0.7% stereotactic-guided biopsy/day	US-guided biopsy: normal (0.75, 0.15); Stereotactic-guided biopsy: normal (1.25, 0.25)

Notes: Imaging equipment is shared resources across all imaging examinations and procedures (e.g., mammography equipment can be used for screening mammography, diagnostic mammography, or stereotactic-guided biopsy). **Abbreviations**: LVN: licensed vocational nurse; AI: artificial intelligence; 2D: two-dimensional; 3D: three-dimensional; mammo: mammography; US: ultrasound; NA: not applicable.

3. Verification/validation

We parameterized the DES model using values that match those observed at our center and using the estimated length of stay mentioned in **Table 1**. The daily average number of patients (i.e., screening + diagnostic + biopsy) from the baseline simulation model was close to 48, matching the observed average number of patients across three weekdays in the center.

4. Analysis

We investigated the following three scenarios to understand the potential impacts of introducing the same-day diagnostic imaging workup and ways to mitigate these changes.

- What is the impact on patient volume, length of stay, and waiting times if the same-day diagnostic imaging workup is introduced with no change to resources in the breast imaging center?
- How does the impact change when more resources (i.e., additional equipment, technologist(s), or hours of operation) are added to support the revised workflow?
- What is the impact of varying the AI algorithm sensitivity?

The results of each scenario were compared to the baseline workflow model. A sensitivity analysis (see **Figure 1-B**) was conducted in the revised workflow model to estimate the effects of varying sensitivity of the AI algorithm on the number of false positives and radiologists' time spent on false positives. A false positive was defined as an abnormal result (i.e., BI-RADS 0, additional evaluation needed) from the AI algorithm, but the radiologist determined that the individual did not warrant a diagnostic examination.

Statistical analysis and implementation details

We ran the baseline and revised DES models with 500 random seeds (i.e., simulated 500 independent clinic days) and averaged the results. Shapiro-Wilk test for normality was used to inform the choice for statistical tests. A two-sample t-test or Wilcoxon rank sum test was used to compare two independent group means, while a paired t-test or Wilcoxon signed-rank test was used to compare two matched group means. In addition, a one-sample t-test was used to compare

the mean of a sample to a known population mean. Two-sided P < .05 was considered significant. Python version 3.7.3 (Python Software Foundation) and the 'SimPy' package were used for analysis. SimPy is a Python-based DES framework enabling rapid development of simulations by employing concepts of processes, shared resources, and events tracked within the simulation environment. (23) (24) The models were developed following the parameters outlined in **Table 1**. The main components of the models are 1) a patient arrival process to generate patients according to interarrival time, 2) a class to store input parameters related to shared resources and process methods for processing times at each step, and 3) a function to specify the sequence of steps followed by patients. Shared resources are noted in the 'capacity' column in **Table 1**.

Results
Validation of the baseline model
As shown in **Table 1**, the center operates nine hours per day for breast imaging and procedures, with 8.5 hours for patient check-in and approximately six patients arriving every hour. There are three check-in staff, one personnel for consenting biopsy patients, three mammography units with three mammography technologists, and two ultrasound units with two ultrasound technologists. Three to four breast radiologists, composed of one to two trainees (resident or fellow) and two breast imaging fellowship-trained attending radiologists, staff the center daily. The public and gowned waiting rooms can accommodate up to twenty and five patients, respectively. There are three changing rooms in the dressing area. Using three-day scheduling data, the estimated proportions of different types of patients each day were 50.3% screening mammography, 45.0% diagnostic imaging, 4.0% ultrasound-guided biopsy, and 0.7% stereotactic-guided biopsy, respectively. The estimated proportions of diagnostic mammography and diagnostic ultrasound, diagnostic mammography only, and diagnostic ultrasound only patients were 80%, 10%, and 10%, respectively. The baseline simulation model using these parameters achieved a mean daily patient volume of 48 patients, the same as the average daily volume estimated using historical scheduling data (p=0.31). Despite attaining a similar patient volume, one discrepancy between the baseline simulation model and clinical practice was noted, where the simulation model over-generated a handful of patients. On average, 52 patients checked in within 8.5 hours in the baseline model, meaning an average of four patients could not complete their examinations with the center operating nine hours a day.

Patient volume
On average, both models generated 52 patients daily (baseline vs. revised, p=0.96). During the nine operating hours a day, 4% fewer patients could complete their scheduled appointments in the revised workflow compared to the baseline workflow (46 vs. 48, p<0.05). Scenarios using added resources were simulated. Adding a new mammography unit and technologist to the revised workflow led to a modest increase in patient volume; however, the number of patients per day remained significantly lower than the baseline workflow (46 vs. 48, p<0.05). Adding a new ultrasound unit and technologist or adding 0.5 operation hours achieved a comparable patient volume as the baseline (p>0.05). The revised workflow achieved greater patient volume by adding 1 hour of operating time compared to the baseline (50 vs. 48, p<0.05).

Table 2. The average number of patients per day (mean [sd]).

Resource	Category	Baseline	Revised	p-value
Equipment & technologists (operating hours=9)	3 mammo + 2 US (BKWIC)	47.7 (6.3)	45.8 (6.0)	<0.01
	3 mammo + 2 US + 1 new mammo	NA	46.3 (6.4)	<0.01
	3 mammo + 2 US + 1 new US	NA	47.1 (6.5)	0.09
Operating hours (3 mammo + 2 US)	9 (BKWIC)	47.7 (6.3)	45.8 (6.0)	<0.01
	9.5	NA	48.5 (6.1)	0.06
	10	NA	50.2 (6.2)	<0.01

Notes: Wilcoxon singed-rank test or paired t-test p-values were derived by comparing each scenario to the baseline. **Abbreviations**: sd: standard deviation; mammo: mammography; US: ultrasound; BKWIC: Barbara Kort Women's Imaging Center; NA: not applicable.

Patient length of stay
Under the center's current operating parameters, the revised workflow had a 24% increase in the length of stay in the imaging center for all types of patients (**Table 3**, 1.26 vs. 1.02, p<0.05), except for stereotactic-guided biopsy patients (2.28 vs. 2.31, p>0.05). The added resource scenarios follow. After adding a new mammography unit and technologist, the length of stay significantly reduced by 0.06 and 0.08 hours compared to the baseline workflow for diagnostic mammography and stereotactic-guided biopsy patients, respectively (p<0.05). After adding a new ultrasound unit and technologist, the length of stay significantly reduced in three patient groups as opposed to the baseline workflow

(p<0.05), which was 0.15 hours for diagnostic mammography and diagnostic ultrasound patients, 0.16 hours for diagnostic ultrasound only patients, and 0.15 hours for ultrasound-guided biopsy patients. Extending the operating hours did not help shorten the length of stay across any patient type.

Table 3. The average length of stay in the imaging center by patient type (hour, mean [sd]).

Patient type	Operating hours=9 3 mammo + 2 US (BKWIC)			Equipment & technologists (operating hours=9)					Operating hours (3 mammo + 2 US)			
				3 mammo + 2 US + 1 new mammo		3 mammo + 2 US + 1 new US			9.5		10	
	Baseline	Revised	p	Revised	p	Revised	p		Revised	p	Revised	p
All patients	1.02 (0.59)	1.26 (0.59)	<0.01	1.15 (0.54)	<0.01	1.12 (0.43)	<0.01		1.21 (0.55)	<0.01	1.23 (0.56)	<0.01
Screening mammo	0.55 (0.17)	0.82 (0.19)	<0.01	0.74 (0.10)	<0.01	0.81 (0.19)	<0.01		0.81 (0.19)	<0.01	0.81 (0.20)	<001
Diagnostic mammo + US	1.56 (0.40)	1.63 (0.46)	<0.01	1.56 (0.40)	0.72	1.41 (0.25)	<0.01		1.59 (0.42)	<0.01	1.60 (0.43)	<0.01
Diagnostic mammo	0.88 (0.19)	0.90 (0.21)	0.03	0.82 (0.13)	<0.01	0.89 (0.20)	0.04		0.89 (0.20)	0.03	0.90 (0.21)	0.03
Diagnostic US	0.98 (0.32)	1.03 (0.35)	<0.01	1.01 (0.34)	0.01	0.82 (0.13)	<0.01		1.01 (0.31)	<0.01	1.01 (0.31)	<0.01
US-guided biopsy	1.99 (0.40)	2.06 (0.43)	<0.01	1.97 (0.41)	0.37	1.84 (0.28)	<0.01		2.03 (0.41)	<0.01	2.04 (0.41)	<0.01
Stereotactic-guided biopsy	2.31 (0.31)	2.28 (0.32)	0.32	2.23 (0.29)	0.02	2.34 (0.29)	0.4		2.26 (0.31)	0.12	2.26 (0.31)	0.14
Combined patient types generated by same-day diagnostic evaluation												
Screening mammo + diagnostic mammo & US	NA	2.18 (0.60)	NA	2.01 (0.44)	NA	1.92 (0.32)	NA		2.12 (0.49)	NA	2.12 (0.51)	NA
Screening mammo + diagnostic mammo	NA	1.41 (0.31)	NA	1.27 (0.19)	NA	1.39 (0.28)	NA		1.41 (0.31)	NA	1.41 (0.31)	NA
Screening mammo + diagnostic US	NA	1.54 (0.45)	NA	1.45 (0.40)	NA	1.31 (0.20)	NA		1.50 (0.40)	NA	1.51 (0.42)	NA

Notes: Wilcoxon rank sum or two-sample t-test p-values were derived by comparing each scenario to the baseline. **Abbreviations**: sd: standard deviation; mammo: mammography; US: ultrasound; BKWIC: Barbara Kort Women's Imaging Center; NA: not applicable.

Patient waiting times
Table 4 shows minimal waiting time at several steps for both the baseline and revised workflow models, including check-in, public waiting room, consent room, and gowned waiting room. For both models, a waiting time of 0.17 hours was observed before entering the dressing room to change into the gown. Under the current center's setting, the revised workflow had longer waiting times before all examinations and procedures (p<0.05) except stereotactic-guided biopsy (p>0.05), and the added waiting time was at most 0.05 hours. Added resources resulted in the following observations. After adding a new mammography unit and technologist, the waiting times before examinations and procedures involving a mammography unit decreased by up to 0.06 hours relative to the baseline workflow (p<0.05) at the cost of increasing up to 0.08 hour waiting times before examinations and procedures involving an ultrasound unit (p<0.05). Similarly, when a new ultrasound unit and technologist were added, the waiting times before ultrasound-related examinations and procedures were reduced by up to 0.14 hours (p<0.05), but this was accompanied by an increase in waiting times for mammography-related examinations or procedures of up to 0.03 hours (p<0.05). Longer operating hours were unhelpful in shortening the patient waiting times (p<0.05).

AI-flagged false positives
As the sensitivity of the AI algorithm increased (i.e., AI recall rate from 10% to 30%), the average number of AI-flagged false positives per day increased (**Table 5**, radiologists agree/disagree%: 70/30%: from 0.7 to 2.3; 80/20%: from 0.5 to 1.5; 90/10%: from 0.2 to 0.7). Accordingly, the average time radiologists spent on AI-flagged false positives per day increased (radiologists agree/disagree% 70/30%: from 0.14 to 0.46 hours; 80/20%: 0.09 to 0.30 hours; 90/10%: 0.05 to 0.14 hours). When the AI recall rate was set to 30%, and the radiologist confirmation rate was set to 70% (i.e., the maximum number of AI-flagged positive examinations combined with the maximum percent of radiologist-deemed negative examinations), the simulated recall rate was 17.5%, the average number of false positive patients per day was two to three, and the average time radiologists spent on false positive examinations was 0.46 hours

Table 4. The average patient waiting time (hour, mean [sd]).

| Waiting for | Operating hours=9 3 mammo + 2 US (BKWIC) | | | Equipment & technologists (operating hours=9) | | | | | Operating hours (3 mammo + 2 US) | | | |
| | | | | 3 mammo + 2 US + 1 new mammo | | 3 mammo + 2 US + 1 new US | | | 9.5 | | 10 | |
	Baseline	Revised	p	Revised	p	Revised	p	Revised	p	Revised	p
Check-in	<0.01 (<0.01)	<0.01 (<0.01)	0.99	<0.01 (<0.01)	0.92	<0.01 (<0.01)	0.92	<0.01 (<0.01)	1	<0.01 (<0.01)	1
Public waiting room	<0.01 (<0.01)	<0.01 (<0.01)	1	<0.01 (<0.01)	1	<0.01 (<0.01)	1	<0.01 (<0.01)	1	<0.01 (<0.01)	1
Consent room	<0.01 (0.02)	<0.01 (0.02)	0.31	<0.01 (0.02)	0.81	<0.01 (0.02)	0.6	<0.01 (0.02)	0.67	<0.01 (0.02)	0.68
Dressing room	0.17 (0.04)	0.17 (0.04)	0.95	0.17 (0.04)	0.73	0.17 (0.04)	0.13	0.17 (0.04)	0.99	0.17 (0.04)	0.83
Gowned waiting room	<0.01 (<0.01)	<0.01 (<0.01)	1	<0.01 (<0.01)	1	<0.01 (<0.01)	1	<0.01 (<0.01)	1	<0.01 (<0.01)	1
Mammo unit (screening mammo)	0.07 (0.16)	0.09 (0.17)	<0.01	0.02 (0.06)	<0.01	0.09 (0.18)	<0.01	0.09 (0.17)	<0.01	0.10 (0.18)	<0.01
Mammo unit (diagnostic mammo)	0.07 (0.15)	0.09 (0.17)	<0.01	0.02 (0.06)	<0.01	0.10 (0.17)	<0.01	0.10 (0.17)	<0.01	0.10 (0.18)	<0.01
Mammo unit (before the US)	0.07 (0.17)	0.09 (0.16)	<0.01	0.02 (0.06)	<0.01	0.10 (0.18)	<0.01	0.09 (0.16)	<0.01	0.09 (0.17)	<0.01
US unit (diagnostic US)	0.17 (0.27)	0.20 (0.29)	<0.01	0.22 (0.32)	<0.01	0.03 (0.08)	<0.01	0.20 (0.29)	<0.01	0.21 (0.30)	<0.01
US unit (after diagnostic mammo)	0.16 (0.26)	0.21 (0.30)	<0.01	0.24 (0.08)	<0.01	0.02 (0.07)	<0.01	0.21 (0.31)	<0.01	0.22 (0.32)	<0.01
US unit (US-guided biopsy)	0.17 (0.28)	0.20 (0.28)	0.02	0.23 (0.34)	<0.01	0.03 (0.09)	<0.01	0.20 (0.29)	<0.01	0.22 (0.31)	<0.01
Mammo unit (stereotactic-guided biopsy)	0.08 (0.15)	0.07 (0.15)	0.33	0.03 (0.09)	<0.01	0.09 (0.18)	0.74	0.07 (0.15)	0.49	0.07 (0.15)	0.56
Mammo unit (after US-guided biopsy)	0.08 (0.17)	0.09 (0.17)	0.03	0.02 (0.08)	<0.01	0.10 (0.18)	<0.01	0.10 (0.17)	0.02	0.10 (0.18)	0.05
Mammo unit (diagnostic mammo after screening mammo) *	NA	0.09 (0.15)	NA	0.03 (0.35)	NA	0.09 (0.15)	NA	0.09 (0.16)	NA	0.09 (0.16)	NA
Mammo unit (diagnostic mammo before US after screening mammo) *	NA	0.09 (0.17)	NA	0.02 (0.07)	NA	0.09 (0.16)	NA	0.09 (0.17)	NA	0.09 (0.17)	NA
US unit (diagnostic US after screening mammo) *	NA	0.20 (0.25)	NA	0.25 (0.35)	NA	0.03 (0.09)	NA	0.21 (0.30)	NA	0.21 (0.30)	NA
US unit (diagnostic US after screening mammo and diagnostic mammo) *	NA	0.21 (0.30)	NA	0.26 (0.36)	NA	0.02 (0.07)	NA	0.22 (0.31)	NA	0.22 (0.32)	NA
Dressing room after examination	<0.01 (<0.01)	0.09 (0.15)	<0.01	0.05 (0.25)	<0.01	0.04 (0.21)	<0.01	0.05 (0.24)	<0.01	0.05 (0.25)	

Notes: * Same-day imaging examinations following a screening mammogram. Wilcoxon rank sum or two-sample t-test p-values were derived by comparing each scenario to the baseline.
Abbreviations: sd: standard deviation; mammo: mammography; US: ultrasound; BKWIC: Barbara Kort Women's Imaging Center; NA: not applicable.

Table 5. Sensitivity analyses for the revised workflow (running hours=9, 3 mammo + 2 US).

Model parameter			Simulation result					
AI positive (%)	Radiologists agree with AI (%)/ disagree with AI (%)	Screeners undergoing same-day diagnostic imaging (%)	AI positive (%)	Radiologists agree with AI (%)	Screeners undergoing same-day diagnostic imaging (%)	No. screeners undergoing same-day diagnostic imaging/day	Average no. false positives (AI positive, radiologist negative)/day	Average radiologists' time spent on false positives/day (hour)
10.0	70/30	7.0	8.9	67.8	6.1	1.5	0.7	0.14
10.0	80/20	8.0	8.6	78.5	6.7	1.6	0.5	0.09
10.0	90/10	9.0	8.6	88.5	7.6	1.9	0.2	0.05
20.0	70/30	14.0	18.1	66.3	12.0	2.9	1.5	0.29
20.0	80/20	16.0	17.4	76.8	13.3	3.2	1.0	0.19
20.0	90/10	18.0	16.8	88.9	14.9	3.6	0.5	0.09
30.0	70/30	21.0	27.0	65.0	17.5	4.3	2.3	0.46
30.0	80/20	24.0	26.4	76.3	20.1	4.9	1.5	0.30
30.0	90/10	27.0	26.0	88.4	23.0	5.5	0.7	0.14

Abbreviations: mammo: mammography; US: ultrasound; AI: artificial intelligence.

Discussion

We designed an AI-aided same-day diagnostic examination workflow for breast screening patients and assessed the potential impact of this workflow on patient experience by comparing it to the baseline. Under the assumptions of the baseline workflow model, introducing the same-day diagnostic workup could lead to a 4% decrease in average daily total patient volume. With no changes to the radiology workflow, this implies adding another 0.5 operating hours to maintain capacity. This has several challenges in practice, such as extra costs associated with longer working hours for all personnel and longer operating hours for the imaging center. Additionally, patients in the revised workflow had longer lengths of stay regardless of examination or procedure type because of the added AI assessment time for screeners and longer waiting time before an examination or a procedure for all patients. Adding a new imaging unit and technologist would be more effective in reducing patient length of stay and waiting times than extending the operating hours of the imaging center if the current radiology workflow is unchanged.

In the worst-case scenario of our simulation, we show that radiologists would spend 0.46 hours on false positives in the AI-aided workflow when 17.5% of screeners need same-day diagnostic imaging examinations. In addition, when defining false positives, we used AI agreement with radiologist interpretation as the reference rather than histopathology diagnosis. However, the false positives in breast cancer screening mentioned in the literature often use the histopathological diagnosis as the reference. Therefore, we will obtain historical breast cancer diagnosis data to estimate the false positive rate based on the histopathological results and incorporate this module into our simulation in future studies. The effect of anchoring bias is essential to examine in radiologist-focused modeling, which this study does not address. Specifically, when radiologists use AI for medical image interpretation, the initial result highlighted by the AI might influence the subsequent review, regardless of its accuracy or relevance. This research is underway.

Aside from the limiting assumptions of holding clinical workflow constant, the simulation models have limitations. First, DES estimates population effects, not individual outcomes. (25) The means used in DES lack the information required to represent a specific individual, including known and unknown factors. Second, the total number of examinations may remain the same over the long term because the diagnostic imaging examinations scheduled the same day as the abnormal screening mammogram are typically arranged shortly after that. The models only simulated the estimated number of appointment slots unavailable when the same-day diagnostic workup is in operation. Third, not all patients who checked in within 8.5 hours could complete their scheduled appointments in the baseline simulation model. In practice, this could be avoided by determining the appropriate check-in time for different types of patients. For example, biopsy patients should check in at least two hours before the center closes. The revised workflow model was also affected by this discrepancy. In the revised workflow, the simulated recall rate was 10% when using the same parameters as the baseline workflow model, which falls within the 5-12% recommended rate by the ACR. (5) However, some patients who needed diagnostic workups after a screening mammogram were unable to complete their appointment within nine hours of operating time. A potential takeaway from this observation is that we would want to schedule screening patients early to ensure sufficient time for completing the same-day diagnostic examinations.

Additionally, this simulation excluded some care processes that an imaging center would include in its overall workflow. The breast imaging center in this study routinely performs screening ultrasound, magnetic resonance imaging (MRI), and MRI-guided biopsies, which were not taken into account in our workflow analysis. Although these imaging examinations and procedures are not directly involved in the proposed workflow, we likely undercounted the patient waiting times and length of stay due to sharing specific resources. For instance, the ultrasound unit is used for both screening and diagnostic patients. As the study assumed that radiologists were not bottlenecks in the workflow, subsequent models can change these assumptions. It may be that with current workflows, introducing the same-day diagnostic workup might divert radiologists from their regular duties, such as interpretation of other breast imaging studies (e.g., screening, MRI examinations) and performance of procedures, with additional effects that this study did not capture. Consequently, this study may have underestimated the impact of same-day diagnostic workup on patient volume and experience. We have identified several workflow questions and potential innovations that can advance future modeling and, ultimately, real-world testing. For example, in the baseline workflow, radiologists experience frequent disruptions, including the need for repeated logins and logouts across five systems: the picture archiving and communication system, mammogram reporting system, dictation system, electronic medical record, and email. They also face interruptions from technologists and trainees seeking guidance on questions and case reviews. In addition to assessing the impacts of same-day diagnostic workup on radiologists' workflow, future studies should prioritize the resolution of these inefficiencies.

Notably, the current models only captured some variations in the system, given the complexity and the number of different tasks of the system. Process capability analysis using past and ongoing time series data can complement simulation modeling to understand the impact of natural variation in the center that was the focus of this research and in any center that implements an AI real-time protocol. The simulation models output 0.17 hours of waiting time before entering the dressing room to change into a gown. The research team did not observe waiting time in the imaging center during clinic shadowing; the patient generator needs to be refined by analyzing more historical data. Overall, like all models, DES models are limited by the assumptions made in the model. The current models can be improved by estimating model parameters and variable distributions from more scheduling and time-motion data.

In future studies, we aim to add internal cost analysis and assess confirmation bias between the AI algorithm and the radiologist. We will pose additional what-if questions, such as (1) what would be the impact on a breast radiologist's productivity (relative value units) if they only interpreted the screening mammograms and same-day diagnostic examinations and (2) what would be the impact on patient throughput if one mammography room and one ultrasound room of the imaging center were dedicated to screening and same-day diagnostic workup. We will conduct a comprehensive cost-benefit analysis associated with implementing this workflow. The costs may include expenses related to acquiring and maintaining AI technology, training personnel, and introducing new imaging equipment. Potential benefits include reduced recalls after screening mammograms, improved diagnostic efficiency leading to earlier detection of breast cancer, and increased patient satisfaction due to shorter waiting times. The DES models built in this work can be reused to mimic the patient flow in other breast imaging centers at our institutions by making minor modifications. As such, the next step is to run simulations at all breast imaging sites selected to pilot the new workflow. Once the pilot studies are conducted, we can use the collected pilot data to, in turn, validate our simulation models. These results will serve as the basis for helping administrative and operational leaders make strategic decisions.

Conclusion
Using DES, we show that introducing an AI-aided same-day diagnostic workup module after an abnormal screening mammogram may reduce daily patient volume but increase the patient's length of stay in the imaging center and waiting times before specific examinations. Potential ways to mitigate these changes include extending the operating hours of the imaging center or purchasing new imaging equipment alongside hiring new technologists. These findings provide the basis for future studies that focus on determining whether the same-day diagnostic workup improves patient satisfaction, decreases patient anxiety, reduces recall rate, reduces time to cancer diagnosis, and improves the efficiency of the breast screening program.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12-49.
2. Njor S, Nystrom L, Moss S, Paci E, Broeders M, Segnan N, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen. 2012;19 Suppl 1:33-41.
3. Swedish Organised Service Screening Evaluation G. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev. 2006;15(1):45-51.
4. Tabar L, Vitak B, Chen TH, Yen AM, Cohen A, Tot T, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology. 2011;260(3):658-63.
5. D'Orsi CJ SE, Mendelson EB, Morris EA, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology. 2013.
6. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med. 2011;155(8):481-92.
7. Ho TH, Bissell MCS, Kerlikowske K, Hubbard RA, Sprague BL, Lee CI, et al. Cumulative Probability of False-Positive Results After 10 Years of Screening With Digital Breast Tomosynthesis vs Digital Mammography. JAMA Netw Open. 2022;5(3):e222440.
8. Kim YS, Jang MJ, Lee SH, Kim SY, Ha SM, Kwon BR, et al. Use of Artificial Intelligence for Reducing Unnecessary Recalls at Screening Mammography: A Simulation Study. Korean J Radiol. 2022;23(12):1241-59.
9. Gilmartin J, Wright K. Day surgery: patients' felt abandoned during the preoperative wait. J Clin Nurs. 2008;17(18):2418-25.
10. Pai VR, Rebner M. How to Minimize Patient Anxiety From Screening Mammography. J Breast Imaging. 2021;3(5):603-6.
11. Lee J, Hardesty LA, Kunzler NM, Rosenkrantz AB. Direct Interactive Public Education by Breast Radiologists About Screening Mammography: Impact on Anxiety and Empowerment. J Am Coll Radiol. 2016;13(1):12-20.
12. Nguyen DL, Oluyemi E, Myers KS, Harvey SC, Mullen LA, Ambinder EB. Impact of Telephone Communication on Patient Adherence With Follow-Up Recommendations After an Abnormal Screening Mammogram. J Am Coll Radiol. 2020;17(9):1139-48.
13. Raza S, Rosen MP, Chorny K, Mehta TS, Hulka CA, Baum JK. Patient expectations and costs of immediate reporting of screening mammography: talk isn't cheap. AJR Am J Roentgenol. 2001;177(3):579-83.
14. Wilson TE, Wallace C, Roubidoux MA, Sonnad SS, Crowe DJ, Helvie MA. Patient satisfaction with screening mammography: online vs off-line interpretation. Acad Radiol. 1998;5(11):771-8.
15. Yoon JH SF, Baltzer PA, Conant EF, Gilbert FJ, Lehman CD, et al. Standalone AI for breast cancer detection at screening digital mammography and Digital Breast Tomosynthesis: A systematic review and meta-analysis. Radiology. 2023;307(5).
16. Vazquez-Serrano JI, Peimbert-Garcia RE, Cardenas-Barron LE. Discrete-Event Simulation Modeling in Healthcare: A Comprehensive Review. Int J Environ Res Public Health. 2021;18(22).
17. Landa P, Sonnessa M, Tanfani E, Testi A. Multiobjective bed management considering emergency and elective patient flows. Int T Oper Res. 2018;25(1):91-110.
18. Hajjarsaraei H, Shirazi B, Rezaeian J. Scenario-based analysis of fast track strategy optimization on emergency department using integrated safety simulation. Safety Sci. 2018;107:9-21.
19. Stahl JE, Roberts MS, Gazelle S. Optimizing management and financial performance of the teaching ambulatory care clinic. J Gen Intern Med. 2003;18(4):266-74.
20. Zhang XG. Application of discrete event simulation in health care: a systematic review. Bmc Health Serv Res. 2018;18.
21. Gunal MM, Pidd M. Discrete event simulation for performance modelling in health care: a review of the literature. J Simul. 2010;4(1):42-51.
22. Pooch UW, Wall JA. Discrete event simulation : a practical approach: CRC Press; 1993.
23. Overview - SimPy 4.0.2 documentation [Available from: https://simpy.readthedocs.io/en/latest/index.html.
24. Kovalchuk SV, Funkner AA, Metsker OG, Yakovlev AN. Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification. J Biomed Inform. 2018;82:128-42.
25. Caro JJ, Moller J. Advantages and disadvantages of discrete-event simulation for health economic analyses. Expert Rev Pharmacoecon Outcomes Res. 2016;16(3):327-9.