Personalized Breast Cancer Screening

Authors names blinded for peer review

Abstract

Abundant literature review and clinical trials indicated that routine cancer screening decrease patient mortality for several common cancer types such as cervical, colorectal, lung and breast cancer. However, current national cancer screening guidelines heavily rely on age as the singular factor of interest and neglect other important medical characteristics of individual patients. This approach either delays screening, which contributes to metastasized tumor growths and increased mortality risk or prescribes excessive screenings, which leads to patient discomfort and financial burden. In this study, we propose a novel machine learning framework to recommend personalized, data-driven, and dynamic screening decisions. We apply this new method to the study of breast cancer mammograms using claims data from 378,840 female patients and show that across different risk populations, personalized screening reduces average delay by 2 to 3 months, with even stronger benefits for individual patients. To evaluate the impact of reduced delay, we develop an approach that estimates the mortality risk. Using this method, we estimate an additional 30% mortality risk reduction using personalized screenings in comparison with the current screening on average.

Keywords: Cancer Screening, Breast Cancer, Personalized Screening, Machine Learning, Early Diagnosis

Highlight

- We consider a new data-driven approach that leverages recent advancements in machine learning and optimization to propose personalized breast cancer screening
- We show that across different risk populations, personalized screening reduced delay by 2 to 3 months, with even stronger benefits for individual patients.
- We estimate that in comparison to the current national guidelines, there is an additional 30% mortality risk reduction using personalized screening approach.
- This method is a generalizable framework that can be readily extended and applied to other cancer screening recommendations.

1 Introduction

Cancer remains the second most common cause of mortality in the United States with an estimated 1.9 million new cases and 608, 570 deaths expected in the United States in 2021 [1, 2]. However, routine cancer screening allows physicians to detect early signs of malicious tumors and propose follow-up treatments to contain tumors from metastasizing beyond the local region [3–5]. Many cancer screenings are proposed to patients, such as low-dose computed tomography for lung cancer, Pap test for cervical cancer, and mammograms for breast cancer. For breast cancer alone, many studies report that routine screening significantly reduced the chance of cancer mortality by about 10–20% in the period of 2010 - 2020 [6–9] and is responsible for saving as many as 27,000 lives in 2018 alone [10].

Yet, despite the proven value of screening, to whom, when, and how often patients should conduct them are subject to lengthy debates. The current United States Preventive Services Task Force (USPSTF) [11], along with many other major institutions such as the American Cancer Society [12], not only have conflicting screening recommendations but also follow a one-size-fits-all

approach with an almost sole focus on age. For breast cancer, the predominant advice is annual screening for women above 50, however with little attention paid to individual characteristics.

In this context, some new studies point to the opportunity to leverage personalized screening recommendations based on individual risks [13–17]. These studies include factors such as family cancer history, and previous biopsies to create a more complete patient profile. Different screening frequencies, as well as screening modalities are also provided as additional options besides the annual screening. However, one significant drawback is their extensive use of disease progression and patient behavior assumptions to constructing simulation environments.

Leveraging recent advancements in machine learning research, this study exploits the potential of using data-driven methods in place of statistical simulations. We propose a new and interpretable framework that uses a state-of-the-art machine learning algorithm, Optimal Survival Trees [18], to personalize screening for breast cancer. Furthermore, as the same patient's risk of developing breast cancer changes over time, we adapt the prescription frequency. The new framework shows a 2 to 3 months reduction of the average cancer detection delay across different risk populations, with even more profound benefits on individual patients. We also build a novel approach to estimate mortality in breast cancer patients that suggests a further 30% mortality risk reduction in comparison to the current one-size-fits-all screening policy on average.

2 Data

In this section, we discuss the data preparation for personalized screening and mortality analysis. First, we retrospectively identify 378,524 female patients profiles from the an insurance company's claims database during the period spanning from December 31, 2009 to January 1, 2020. Then, to focus the study on female population, we exclude male or mixed gendered patients, leaving 349,088 patients. Finally, we exclude patients who have neither a breast cancer history, nor prior screenings, resulting in a total of 329,156 remaining patients.

2.1 Personalized Screening

To study how to personalize screenings to prevent first-time breast cancer occurrence, we focus only on claims prior to the first breast cancer diagnosis of each patient. Another exclusion criterion involves the unusual challenge faced by the healthcare system under significant disruptions caused by the COVID-19 pandemic. Patients were unable to visit their physicians regularly due to the scarcity of healthcare workers. Thus, we exclude all claim transactions after January 1st, 2020 in the study.

In this study, a number of important clinical and demographic factors that suggest potential links to breast cancer risk were identified and selected based on past literature. These factors encompass a wide range of patient medical profiles, including other disease diagnoses relating to increased breast cancer risk, common comorbidities, drugs prescribed, as well as procedures conducted [19]. The objective of this study relies only on factors that can be observed by patients themselves or during routine physician visits, as opposed to those that require deeper, more sophisticated and time-consuming diagnostics, such as breast density measurements. We used Standardized International Classification of Diseases (ICD) version 10 code, Current Procedural Terminology (CPT) code, National Drug Code (NDC), as well as employee demographic information to identify these factors of interest as input. Table 1 depicts the list of all features in the analysis.

The primary outcome of the study is the delay between the onset of breast cancer's screening detectable phase and an abnormal screening mammogram that signals breast cancer. However, one drawback of claims data is its lack of clear indication of the results of medical screenings, making it challenging to decide the precise timestamp of screening detectable breast cancer. Furthermore, for patients who detected cancer through other means besides screening, it is possible that no abnormal screening mammogram exists. Instead, we propose the notion of a **pessimistic delay**, and define it as the length of time between the last indication of a cancer-free period and the first indication of breast cancer onset. The pessimistic delay represents an upper bound on the time of

cancer onset as illustrated in Figure 1. Depending on the patients claims, these timestamps are defined differently. We give the following examples as an illustration of how to calculate the pessimistic delay:

- A patient who has only one screening claim at day 100 and her first breast cancer claim at day 300: In this case, the last indication of a cancer-free period is 100, and the first indication of breast cancer onset is 300. The pessimistic delay is 200.
- A patient who has two screenings, one at day 100, the other at day 270, and her first breast cancer claim is at day 300: In this case, the class indication of cancer-free period is again 100, but the first indication of breast cancer onset is 270. The pessimistic delay is 170.

This worst-case estimation is illustrated in Figure 1. It is worth noting that both in-situ, as well as malignant breast cancer claims, are considered an indication of breast cancer onset.

2.2 Mortality Analysis

A key measure of success of any screening guideline is the number of deaths averted, or the number of lives saved. To study the effect of personalized screening over a one-size-fits-all screening, we perform an additional analysis to evaluate the effect of pessimistic delay on the mortality rate of patients with at least one breast cancer claim. As the first step, this side study restricts the population to 8,219 patients with at least one breast cancer claim only. Secondly, to ensure that patients selected in the mortality study exited the system instead of merely not returning to their physicians for follow-up, we apply an additional criterion that the patient exited before June 2019, resulting in a final total count of 1,544 patients.

We select features to reflect severe conditions linked to mortality using CPT, ICD and NDC drug codes. Additionally, we calculate pessimistic delay (in months) using patient screening history and add it as an additional feature. Referring again to Figure 1, note that some first screening mammograms occur after breast cancer onset. In this scenario, pessimistic delays could not be calculated since they do not satisfy our definition of the pessimistic delay. Thus, these patients are removed from the study.

Inferring mortality is an especially challenging task in claims data due to the lack of precise knowledge of each claim transaction taken. It is possible that doctors or hospitals may suggest additional services recorded outside of the claims system after their exits. To avoid these issues in the study, we define patients as deceased in the following scenarios. First, we include patients who exited the claims system with a discharge location of ambulance, emergency room, hospice, end-stage renal, or intensive care unit. Secondly, we also consider as deceased, patients with at least two consecutive days in the emergency room during their last 30 days of claim history. This method allows us to pin point the exact reasoning behind why a certain patient exits the system, and is an important aspect of the analysis.

3 Methods

In this section, we discuss our methodologies for personalized screening. The important first step involves generating an individualized, probabilistic survival projection that allows us to determine the trajectory of a patient's cancer risk. The second step utilizes a survival curve to make an optimal trade-off between pessimistic delay and over-screening. Lastly, we find an interpretable way to translate the benefit of reduced pessimistic delay to reduced breast cancer mortality risk.

3.1 Obtaining Survival Risk Progression

In the first step of model development, given a patient's past medical history, we first aim to predict a patient's breast cancer survival. For this task, we use Optimal Survival Trees [18], a novel, tree-based machine learning survival model. Optimal Survival Trees stratify the patient population into subgroups by following decision paths. For example, if the tree decides that patients older than 50 have significantly different survival trajectories than those younger than 50, then these will be classified into two distinct subgroups. However, Optimal Survival Trees is not the only method available for accurate survival predictions, an example of another popular method is random survival forest [20]. This is an ensemble tree-based

Table 1 Features used in study, their data as well as feature types

Feature Name	Feature Type	Data Type
Age	Demographic	Numeric
Smoker	Demographic	Binary
Ovarian Cancer	Diagnosis	Binary
Thyroid cancer	Diagnosis	Binary
Ovary removal	Diagnosis	Binary
Hypothyroidism	Diagnosis	Binary
Hyperthyroidism	Diagnosis	Binary
Hypertension	Diagnosis	Binary
Diabetes	Diagnosis	Binary
Rheumatologic Disease	Diagnosis	Binary
Family history of malignant neoplasm	Diagnosis	Binary
Personal history of malignant neoplasm	Diagnosis	Binary
Overweight and obesity	Diagnosis	Binary
Suspicious findings in breast	Diagnosis	Numeric
Cardiovascular	Drug	Numeric
Endocrine	Drug	Numeric
Anti-infective	Drug	Numeric
Contraceptive	Drug	Numeric
Dermatological	Drug	Numeric
Immunosuppressive	Drug	Numeric
Biopsy	Procedure	Numeric
Ultrasound	Procedure	Numeric
MRI	Procedure	Numeric
BRCA test	Procedure	Numeric
Radiation Therapy	Procedure	Numeric

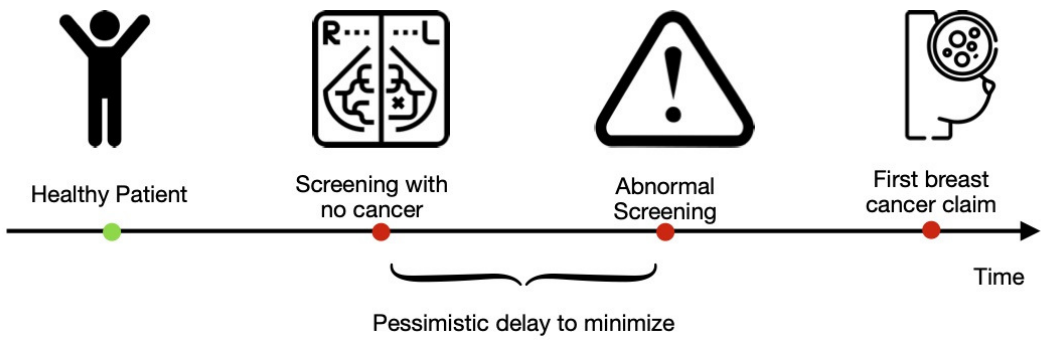

Fig. 1 Pessimistic pessimistic delay is calculated between a definite indication of cancer onset and normal health.

survival model that builds a forest of survival trees and takes consensus from all individual predictors to arrive at a prediction. In general, random survival forest has better performance measures than Optimal Survival Trees, but with the disadvantage of significantly reduced interpretability. For both these models, we evaluate their performance on future unseen data, and train only on a subset of selected patients (80%) and validate model performance on the remaining patients (20%). During training, the model will select hyperparameters to achieve the best predictive performance on the training set. To avoid overfitting, we also apply 5-fold cross-validation in the training procedure.

Table 2 Features used in mortality analysis, their data as well as feature types.

Feature Name	Feature Type	Data Type
Abnormalities of breathing	Diagnosis	Binary
Nausea and vomiting	Diagnosis	Binary
Ascites	Diagnosis	Binary
Lab test for magnesium	Procedure	Binary
Lab test for phosphate	Procedure	Binary
Level IV surgical pathology (biopsy)	Procedure	Binary
Nervous System Drug	Drug	Binary
Pessimistic delay	Calculated	Numeric

3.2 Minimizing Patient pessimistic delay

In the second step, using the projected breast cancer survival curve, we calculate an optimal follow-up screening time. The problem we seek to solve is

$$\arg\min_{h} \mathbb{E}[\max(h - Y, 0)] + \frac{\lambda}{\mathbb{E}[Y \mid Y \leq h]}$$

where

- Y: random variable indicating the time of breast cancer onset,
- h: time until the next screening,
- λ: penalty of screening too early,
- $\mathbb{E}[\max(h - Y, 0)]$: expected pessimistic delay,
- $\mathbb{E}[Y \mid Y \leq h]$: expected conditional time of breast cancer onset given that cancer occurs before screening,
- $\frac{1}{\mathbb{E}[Y|Y \leq h]}$: rate of breast cancer onset given that cancer occurs before screening.

The first term indicates the expected pessimistic delay. Specifically, if h is larger than Y, meaning that cancer onset occurs before screening, then the pessimistic delay is equal to $h - Y$, which is the time between screening and breast cancer onset. If h is smaller than Y, meaning that cancer signs do not appear at the time of screening, then the pessimistic delay is 0. The second term penalizes prescribing screening too early, or over-screening. It captures the rate of breast cancer onset, which is the number of people who will have breast cancer onset given $Y \leq h$ over a long horizon t. While a doctor may recommend a patient to return for an early screening (small h), which effectively reduces the expected pessimistic delay, doing so may increase the rate of cancer onset, and cause over-screening. Thus, by aiming to minimize both terms simultaneously, a patient can decide to gain less frequent screenings but risk longer delay, and vice versa. Instead of a cost-effectiveness analysis, this framework allows the patients, as well as their clinicians, to more intuitively understand the trade-offs of different screening decisions. Note that the framework calculates all screening options on a monthly basis.

3.3 Construct Dynamic Schedules

In the final step, we use the above decisions to propose a dynamic schedule. Each time a patient visits their doctor, we find the patient's future survival risk progression using Optimal Survival Trees. Then, with varying different degrees of over-screening penalty, λ, we solve the optimization problem presented and obtain different optimal screening times. The method only updates an unrealized screening prescription to either sooner or later if the survival risk progression of a patient changes. An example is as follows: a patient comes to a doctor visit and files the corresponding claims on day 1. Using these claims and their medical history, the proposed method suggests they come back on day 300 for a screening mammogram. However, the patient comes back for another doctor visit on day 100 with a newly developed comorbidity, such as a change in drug intake dosage. The algorithm incorporates this new knowledge and updates the screening suggestion. Now, it decides

the patient should come back sooner, in 90 days, on day 190 instead.

3.4 Evaluating Number of Deaths Averted using Mortality Analysis

After obtaining the newly proposed personalized screening schedules, we aim to develop a method to translate the benefit of reduced delay to the benefit of patient mortality. To ensure interpretability, we use a logistic regression model. The model uses pessimistic delay and other patient characteristics to predict patient mortality status at the time of their hospital exit. The coefficients and significance levels of these features imply their impact on patient mortality. Similar to previous experiments, we split the dataset into train (80%) and test (20%) sets. We observe a significant class imbalance, i.e., the cohort of patients who survived is significantly larger in sample size than that of patients who died. Therefore, prior to the training of logistic regression, we apply a class-balancing method, Randomly Over Sampling Examples (ROSE), to generate synthetic samples in the minority group to balance the two subgroups. However, to keep our validation process valid, we do not modify the test set. Finally, we delete all statistically insignificant features and retrain the logistic regression model.

After obtaining the final model, we calculate the difference in mortality risk between current and personalized screening policies. First, we define the following variables

- P_0: probability of breast cancer death with current screening policy,
- P_1: probability of breast cancer death with m-month of reduced pessimistic delay given by personalized screening policy.

Considering patient mortality as the positive outcome, we calculate P_0 using logistic regression as follows

$$P_0 = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}},$$

where $x_1, x_2, \cdots, x_n$ are features of the logistic regression, and $\beta_0, \beta_1, \cdots, \beta_n$ are the coefficients associated with each feature. We assume here that x_n is the pessimistic delay. Given an m-month benefit of pessimistic delay, we can calculate the new probability of patient mortality risk as follows

$$P_1 = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n (x_n - m))}},$$

where

$$e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)} = \frac{1}{P_0} - 1 = \frac{1 - P_0}{P_0},$$

leading to

$$P_1 = \frac{1}{1 + \frac{1 - P_0}{P_0} e^{m\beta_n}}.$$

Lastly, the difference of the two probabilities, $P_1 - P_0$, is the benefit of decreased pessimistic delay on patient mortality risk.

Ethics approval declarations and consent to participate: Not applicable.

4 Results

The following sections detail the numerical results of the proposed methodology. We first evaluate the survival model, Optimal Survival Trees, whose accuracy performance lays the foundation of proposed personalized screening. The next step is to report the difference in pessimistic delay between current and personalized screening using dynamic scheduling. Lastly, we discuss results from logistic regression, which includes insightful information on how pessimistic delay impacts patient mortality risk.

4.1 Predictive Performance of Optimal Survival Trees

To assess the predictive performance of Optimal Survival Trees [18] and random survival forest [20], we calculate traditional survival model evaluation metrics including Harrel's C statics and AUC. Both measure the goodness of fit for models that produce risk scores, where survival analysis generally uses Harrel's C statistics [21], and binary classification analysis uses AUC [22]. Most

medical studies consider Harrell's C statistics and AUC scores of approximately 0.6 - 0.8 as reasonable scores for accuracy. Furthermore, due to the imbalanced nature of our outcome (only 2% breast cancer patients), we also report sensitivity and specificity.

Gail, Tyrer-Cuzick, and BCSC [23–26] are some of the most widely reported and cited risk estimation models for breast cancer. These models use family breast cancer history, age, biopsy histories, and other factors to decide the risk of a patient's future breast cancer occurrence. Respectively, they report 5-year AUC values of 0.61, 0.62, 0.64. However, note that these reported values are only descriptive, as we obtained these values from papers using different data and features. In comparison, the Optimal Survival Trees have higher AUCs than the models above. Below in Table 3 and Table 4 are the train and test Harrel's C statistics and AUC at years 1, 2, 3, 4, and 5. First, we observe that training and testing Harrell's C are only 3% apart, and for AUC the difference is at most 1% in all years. This small difference indicates that the classification is successful and the model achieves a low mismatch rate independent of the year and dataset. Furthermore, it implies the model has good potential to generalize well in future unseen datasets as well.

Table 3 Harrel's C statistics of Optimal Survival Tree in Train and Test sets.

Model	Training Set	Test Set
Optimal Survival Tree	0.673	0.657
Random Survival Forest	0.807	0.720

Figure 2 demonstrates the full tree and magnified as an example. The red solid line in each node is the Kaplan-Meier curve for that sub-population. In addition, the dotted green line represents the survival curve of the population in the lower/left child, and the dotted blue line represents the survival curve of the population in the upper/right child. In the leaf nodes, the dotted black line shows the survival curve of the entire population. The number n in each node represents the number of patients, and expected survival indicates the expected survival day for those patients. In this example, patients with more than 1 MRI scans have a significantly worse survival curve than those who have none or only 1. The tree further splits to two sub-populations, where in the first one, patients with age older than 59 have worse survival. In the second sub-population, patients with overweight or obesity will also have significantly worse survival than those who do not.

4.2 An Example of the Power of Personalized Screening

In this section, we give an example of the significant benefit of using personalized screening. This is a patient who, at the beginning of their claim history, was 48 years old and took very few dosages of anti-infective and endocrine drugs. However, at around day 1060, approximately 3 years later, they began developing another type of cancer. This additional cancer occurrence deteriorates their health significantly, and we, therefore, expect that they are more susceptible to breast cancer. In their original real-life doctor visits, the patient conducted regular annual screenings on days 346, 710 and 1030, and observed an abnormal screening that signaled the occurrence of breast cancer on day 1394. Under this scenario, screening recommendations remained unchanged even after the discovery of a severe risk factor that significantly increased the patient's risk of breast cancer.

However, with personalized screening, we take into account individual medical conditions. When the patient has few comorbidities implying above-normal breast cancer risk at the beginning, the prescription suggests against unnecessarily frequent screening. The algorithm suggests the patient comes in for a screening on days 540 and 1080, both at around 1.5 years, less frequent than the original annual screening guideline. In contrast, as soon as the patient starts having another personal cancer, personalized screening captures a significantly increased risk. This time, the algorithm proposes that the patient returns for mammograms at day 1320 and 1560, both at 8 months instead, which is drastically adapted to be more frequent. The pessimistic delay of breast cancer detection is significantly improved. Following current annual screening guidelines, the patient has a pessimistic delay of 364 days, whereas the personalized pessimistic delay is significantly shorter.

Table 4 AUC, sensitivity and specificity of Optimal Survival Tree at years 1, 2, 3, 4, and 5 in both train and test.

Time Step	Train AUC	Train Sensitivity	Train Specificity	Test AUC	Test Sensitivity	Test Specificity
Year 1	1	1	1	0.677	0.000	1.000
Year 2	1	1	1	0.681	0.004	1.000
Year 3	1	1	1	0.712	0.004	1.000
Year 4	1	1	1	0.738	0.004	0.999
Year 5	1	1	1	0.743	0.004	1.000

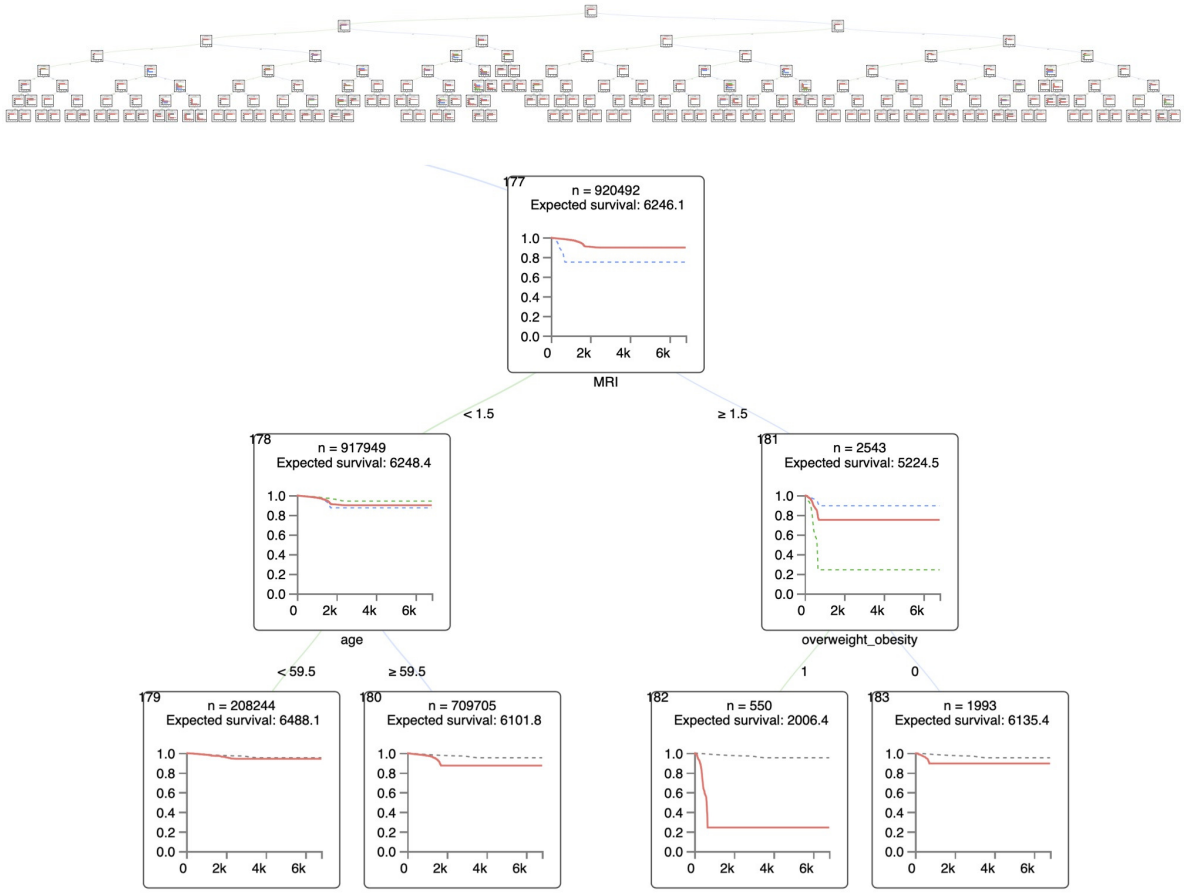

Fig. 2 Part of the full Optimal Survival Tree.

Depending on the mammogram results, we calculate the personalized pessimistic delay as follows. The next remarks are related to Figure 3.

- If prescribed screening #2 is positive, then the pessimistic delay is 50 days.
- If prescribed screening #2 is negative and #3 is positive, then the pessimistic delay is 240 days.
- If prescribed screening #2 and #3 are negative and #4 is positive, then the pessimistic delay is 240 days.

Taking the worst-case pessimistic delay of the above three scenarios, 240 days, personalized screening still reduces the current pessimistic delay by as much as 4 months. Furthermore,

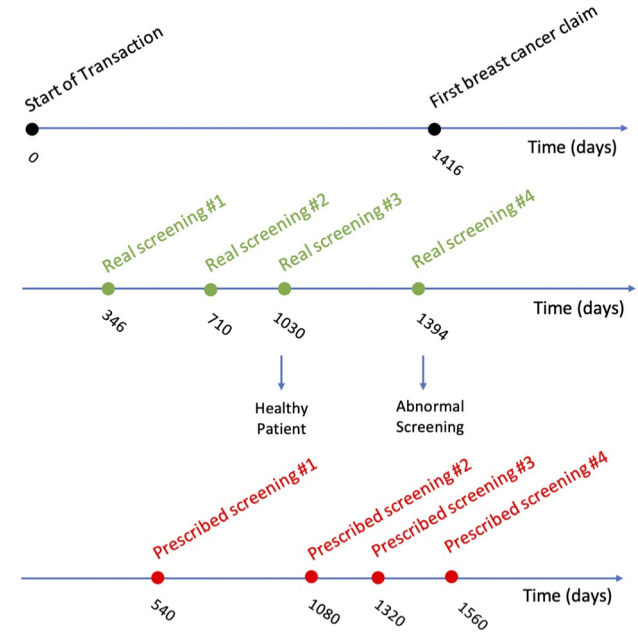

Fig. 3 Timeline of the patient's events of interest, including both current claims data and proposed personalized screening.

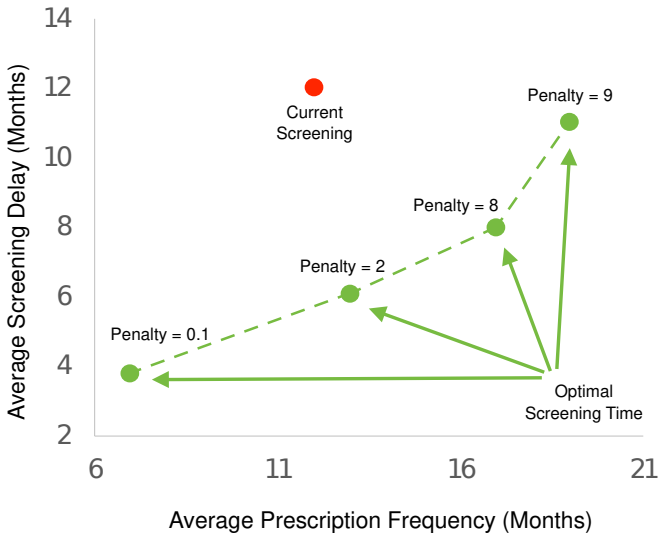

Fig. 4 Efficient Frontier demonstrating the tradeoffs between benefit and inconvinience of different screening decisions.

the patient and their health care provider also have the capability to tradeoff the time for the next screening mammogram based on different costs and benefits by varying the degree of over-screening penalty. As illustrated in Figure 4, the framework can decide the optimal screening time given these different choices. Each of the penalties will result in different prescription frequencies, which leads to different pessimistic delays. For example, a patient can choose a 16-month average interval over a 13-month interval knowing they are risking an additional 1-month pessimistic delay.

4.3 Benefits of Personalized Screening Across the Entire Population

In this section, we evaluate the benefits of personalized screening across the entire population. For a fair comparison, we first ensure that the average time elapsed between personalized screenings is close to that of the real screenings. To achieve this, we calculate the average time interval between each screening mammogram in the original claim. We then tune the over-screening penalty parameter for each patient so that the proposed personalized screening interval matches as closely as possible to that of the original claim.

To study the effects of personalized screening, we stratify patients into three risk groups. To achieve this, we calculate a one-year survival risk for each patient at each claim transaction time by using the survival curves generated from Optimal Survival Trees. We then stratify each claim transaction into high, medium, and low-risk groups by population cutoffs of 24%, 57%, and 19%, respectively. We choose cutoff thresholds such that as risk increases, the average screening interval decreases monotonically. In the next step, we determine a patient's overall risk allocation by the most frequent risk group their claim transactions cumulatively belong to. We exclude outlier patients with average screening intervals of less than 90 days and above 1000 days from the study, which accounts for 2% of the total patient count. Lastly, we report the number of patients in both training and test sets in Table 5.

Table 5 Number of Patients in both train and test sets across different risk groups.

Risk Group	Training Set	Test Set
Highest Risk	440	124
Medium Risk	517	133
Lowest Risk	187	40

Finally, we calculate pessimistic delay for both current and personalized screenings. Figure 5 indicates the performance of personalized screening on filtered patients in both train and test. In both data sets, we observe a significant benefit averaging 2 to 3 months of pessimistic delay reduction across all risk groups. Table 6 demonstrates the detailed numerical results. As patient cancer risk increases, the benefit also increases. The test set performs on average worse than the training set by only less than half a month, which indicates that our model does not over-fit and generalize well on unseen data. We also report the p-value from a t-test comparing the average pessimistic delay between current screening and personalized screening with 95% confidence interval in Table 7, and see that across all risk groups and in both train and test sets, our method significantly outperforms current screening. Lastly, we report the p-value from a t-test comparing the average pessimistic delay between high and medium risk groups as well as medium and lowest risk groups. Respectively, with a 95% confidence interval, we have 0.005 in train and 0.246 in test for the high vs. medium comparison, and 0.008 in train and 0.210 in test for the medium vs. low comparison.

Table 6 The benefit of personalized screening over current screening in number of months across all different risk groups.

Risk Group	Training Set	Test Set
Highest Risk	3.1 months	2.7 months
Medium Risk	2.7 months	2.4 months
Lowest Risk	2.2 months	1.8 months

Table 7 The benefit of personalized screening over current screening is statistically significant across all different risk groups.

Risk Group	Training Set P-value	Test Set P-value
Highest Risk	0	5.0e-223
Medium Risk	0	1.8e-236
Lowest Risk	6.5e–246	9.6e-20

As illustrated by Figure 5, there exists different benefits of personalized screening. The benefit can be as large as 10 months, but also can potentially be a negative impact of 5 months. However, on average, personalized screening offers a clear advantage over a one-size-fits-all approaches. We hypothesize that current guidelines do not address the needs of a large group of high-risk patients who

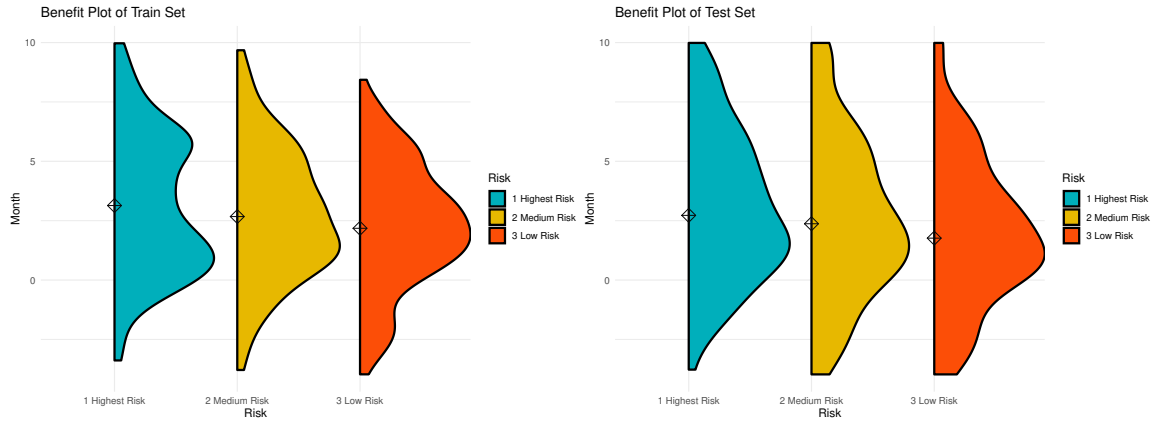

Fig. 5 Results showing on average 2-3 months of pessimistic delay reduction across different risk groups.

should be treated with additional care. The high-risk group thus obtains additional benefits from personalized screening.

4.4 Number of Deaths Averted by Reduced pessimistic delay

The logistic regression for mortality analysis performs strongly with a train AUC of 0.988 and a test AUC of 0.859. Table 8 reports the coefficients of each feature as well as their standard deviation. We also use the Wald test to report the statistical significance of each feature.

In particular, the coefficient of the pessimistic delay is 0.136, which we use in the previously discussed formula to calculate the benefit of reduced pessimistic delay on patient mortality. In Table 9, Current Policy Risk is the probability of breast cancer death with current screening guidelines and Personalized Policy Risk is the probability of breast cancer death with a 3-month reduced pessimistic delay given by personalized screening. We calculate the Percentage Difference, defined as (Personalized Policy Risk - Current Policy Risk)/(Current Policy Risk). By and large, across different mortality risk groups, the benefit of the reduced pessimistic delay is very significant, where in low-risk groups, the benefit could be as large as 30%, and in high-risk groups 5%.

We also calculate the percentage difference on datasets used to train the logistic regression model and report their results in Table 10. Across the board, we observe an average of 30% reduction in mortality risk in comparison with the previously used screening policy.

5 Discussion

In this section, we illustrate the power of personalized screening and discuss the advantages of our selection of claims data, the potential implementation of the method into real-world applications, as well as the limitations in the study.

5.1 Power of Personalized Screening

Patients with different medical characteristics form significantly heterogeneous cohorts. Distinct comorbidities as well as past medical history separate patients into cohorts with different breast cancer risks. By personalizing screenings for these patients, there is a significant reduction of the pessimistic delay of 2 to 3 months on average. We also observe an even greater reduction for individual patients of up to 10 months.

Given the same screening frequency, our advantage lies in the fact that we take into account personalized cancer risks as they change dynamically, and adapt our method to accelerate or decelerate screenings. Figure 6 shows the difference of first screening and last screening interval distribution when a patient is known to develop other personal cancer in between these two screenings. Current screening paradigms see little difference or updates made to adjust patient screening frequencies and largely rely on prescribing annual screenings. Yet, personalized screening shows significant changes in screening recommendation distribution and promotes the avoidance of unnecessary screenings before cancer risk factors appear, and instead accelerates once cancer risk factors appear. In the

Table 8 The coefficients and their statistical significance of all features.

Feature	Coefficient	Std. Error	Z value	P value	Significance
Intercept	-7.21	1.51	-5.11	3.22E-7	0.01E-1
Abnormalities of breathing	3.74	0.95	3.95	7.75E-5	0.01E-1
Nausea and vomiting	4.12	1.18	3.48	0.49E-4	0.01E-1
Ascites	5.21	1.13	4.59	4.37E-6	0.01E-1
Lab test of Magnesium	5.52	1.15	4.82	1.44E-6	0.01E-1
Lab test of Phosphate	3.50	1.02	3.44	5.91E-4	0.01E-1
Level IV surgical pathology	4.64	1.13	4.09	4.31E-5	0.01E-1
Nervous system drug	2.61	0.73	3.56	3.77E-4	0.01E-1
Pessimistic delay	0.136	0.05	2.98	2.87E-3	0.01

Table 9 Summary statistics of the benefit on patient mortality risk using personalized screening over current screening policy.

Current Policy Risk	Personalized Policy Risk	Percentage Difference (%)
0.10	0.069	31
0.20	0.14	29
0.30	0.22	26
0.40	0.31	23
0.50	0.40	20
0.60	0.50	17
0.70	0.61	13
0.80	0.73	9.1
0.90	0.86	4.8

current one-size-fits-all screening approach, the initial screening has an average of a 15 month interval, and ends with a slightly more frequent 14.4 month interval. However, personalized screening starts with a 15.8 month interval, and ends with an 11.1 month interval. This result shows the power of personalized screening by adapting to a patient's constantly changing breast cancer risk. Using the same number of screenings and similar screening intervals, personalized screening better captures potential changes and offers more customized recommendations.

5.2 Advantages of Claims Data

As the foundation of the proposed method, claims data offer the advantage of wide availability as well as a large sample size. This is a key challenge in many other medical studies, where the acquisition of quality data poses a significant obstacle for the development of downstream modeling tasks. In comparison, claims data also provides broad applicability, accessible evaluation as well as patient autonomy in understanding the impact of different screening decisions. Even without detailed knowledge of breast density and other difficult to obtain metrics, the patients can have an intuitive first-hand exposure to personalized screening recommendations using claims data.

5.3 Potential Implementation Pipeline

The proposed algorithm can be easily integrated into many medical application settings by the following pipeline. When a patient first visits a clinic, a doctor will first update the patient's medical information including the details of the current visit. Using this information, the doctor can run the model and instantly learn the next best screening time. This screening time will

Table 10 Percentage Difference of Reduced Mortality Risk with Personalized Screening.

Dataset	Average	Standard Deviation	Maximum	Minimum
Train	0.30	0.090	0.34	4.9E-08
Test	0.31	0.063	0.34	3.6E-05

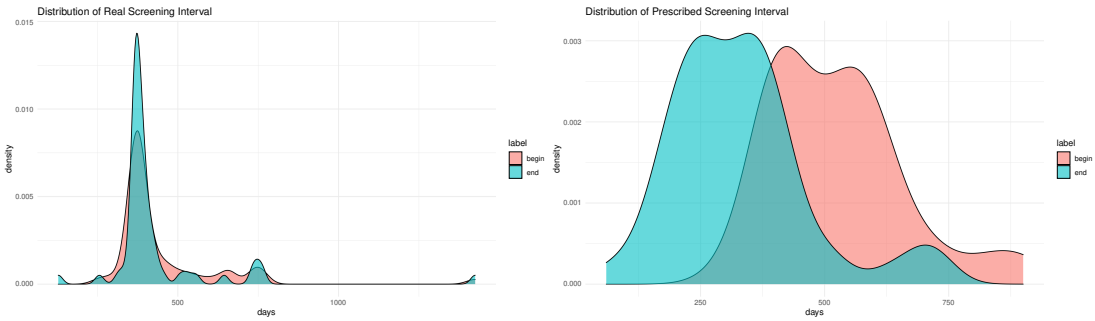

Fig. 6 Personalization of screening is also important in applying dynamic screening for patients.

reflect the patient's changes since their last visit, and will decelerate or accelerate in time or stay the same. Gathering similar information across different clinics and observing the outcomes of the recommendations, we rerun the model periodically, for example, per month, according to the updated data poll to refine the method.

5.4 Limitations

However useful this study may be, we also point to its limitations. First, though claims data provide many advantages as described in the previous section, it is not as detailed as, for example, EHR data. It also does not provide the results of screening or the images obtained in the screenings, both could potentially add significant values to the method. Secondly, we also acknowledge that depending on the length of a patient's claims history, a doctor's recommendations may be biased. Finally, we assume in the results section that patients will adhere to our screening recommendations, which is possibly difficult to enforce in practice.

6 Conclusion

We introduce a personalized, dynamic and data-driven framework to provide personalized cancer screening recommendations. Specifically, we apply this method for women across different characteristics to help reduce delay from breast cancer. By using claims data, this method implies broad applicability and easy extensibility for future usage to hospitals and at-home evaluations. We show that the proposed method significantly reduces breast cancer pessimistic delay by on average 2 to 3 months, and up to 10 months in individual patients. Most importantly, such a reduction of pessimistic delay can save as many as 30% more patients in comparison to current one-size-fits-all guidelines. This novel screening recommendation system offers powerful potentials to help reduce screening detection delays as well as further reducing mortality of cancer patients.

Statements and Declarations

- Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
- Conflict of interest/Competing interests: No funding was received to assist with the preparation of this manuscript.
- Ethics approval: Not applicable
- Consent to participate: Not applicable
- Consent for publication: Not applicable
- Availability of data and materials: Not available
- Code availability: Not available

References

[1] Royce, T.J., Hendrix, L.H., Stokes, W.A., Allen, I.M., Chen, R.C.: Cancer screening rates in individuals with different life expectancies. JAMA Internal Medicine **174**(10), 1558 (2014). https://doi.org/10.1001/jamainternmed.2014.3895

[2] Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2021. CA: A Cancer Journal for Clinicians **71**(1), 7–33 (2021). https://doi.org/10.3322/caac.21654

[3] Nothacker, M., Duda, V., Hahn, M., Warm, M., Degenhardt, F., Madjar, H., Weinbrenner, S., Albert, U.-S.: Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. a systematic review. BMC Cancer **9**(1) (2009). https://doi.org/10.1186/1471-2407-9-335

[4] Leitch, A.M., Dodd, G.D., Costanza, M., Linver, M., Pressman, P., McGinnis, L., Smith, R.A.: American cancer society guidelines for the early detection of breast cancer: update 1997. CA: A Cancer Journal for Clinicians **47**(3), 150–153 (1997). https://doi.org/10.3322/canjclin.47.3.150

[5] Kösters, J.P., Gøtzsche, P.C.: Regular self-examination or clinical examination for early detection of breast cancer. Cochrane Database of Systematic Reviews (2003). https://doi.org/10.1002/14651858.cd003373

[6] Burton, R., Stevenson, C.: Assessment of breast cancer mortality trends associated with mammographic screening and adjuvant therapy from 1986 to 2013 in the state of victoria, australia. JAMA Network Open **3**(6), 208249 (2020). https://doi.org/10.1001/jamanetworkopen.2020.8249

[7] Shapiro, S.: Periodic breast cancer screening in reducing mortality from breast cancer. JAMA: The Journal of the American Medical Association **215**(11), 1777 (1971). https://doi.org/10.1001/jama.1971.03180240027005

[8] Fletcher, S.W., Black, W., Harris, R., Rimer, B.K., Shapiro, S.: Report of the international workshop on screening for breast cancer. JNCI: Journal of the National Cancer Institute **85**(20), 1644–1656 (1993). https://doi.org/10.1093/jnci/85.20.1644

[9] Shapiro, S.: Evidence on screening for breast cancer from a randomized trial. Cancer **39**(6), 2772–2782 (1977). https://doi.org/10.1002/1097-0142(197706)39:6⟨2772::aid-cncr2820390665⟩3.0.co;2-k

[10] Hendrick, R.E., Baker, J.A., Helvie, M.A.: Breast cancer deaths averted over 3 decades. Cancer **125**(9), 1482–1488 (2019). https://doi.org/10.1002/cncr.31954

[11] Breast Cancer: Screening (2016). https://uspreventiveservicestaskforce.org/uspstf/recommendation/breast-cancer-screening

[12] Cancer Screening Guidelines: Detecting Cancer Early. https://www.cancer.org/healthy/find-cancer-early

[13] Schousboe, J.T., Kerlikowske, K., Loh, A., Cummings, S.R.: Personalizing mammography by breast density and other risk factors for breast cancer: Analysis of health benefits and cost-effectiveness. Annals of Internal Medicine **155**(1), 10 (2011). https://doi.org/10.7326/0003-4819-155-1-201107050-00003

[14] Trentham-Dietz, A., Kerlikowske, K., Stout, N.K., Miglioretti, D.L., Schechter, C.B., Ergun, M.A., van den Broek, J.J., Alagoz, O., Sprague, B.L., van Ravesteyn, N.T., Near, A.M., Gangnon, R.E., Hampton, J.M., Chandler, Y., de Koning, H.J., Mandelblatt, J.S., and, A.N.A.T.: Tailoring breast cancer screening intervals by breast density and risk for women aged 50 years or older: Collaborative modeling of screening outcomes. Annals of Internal Medicine **165**(10), 700 (2016). https://doi.org/10.7326/m16-0476

[15] Román, M., Sala, M., Domingo, L., Posso, M., Louro, J., Castells, X.: Personalized breast cancer screening strategies: A systematic review and quality assessment. PLOS ONE **14**(12), 0226352 (2019). https://doi.

org/10.1371/journal.pone.0226352

[16] and, L.J.E.: The WISDOM study: breaking the deadlock in the breast cancer screening debate. npj Breast Cancer **3**(1) (2017). https://doi.org/10.1038/s41523-017-0035-5

[17] Onega, T., Beaber, E.F., Sprague, B.L., Barlow, W.E., Haas, J.S., Tosteson, A.N.A., Schnall, M.D., Armstrong, K., Schapira, M.M., Geller, B., Weaver, D.L., Conant, E.F.: Breast cancer screening in an era of personalized regimens: A conceptual model and national cancer institute initiative for risk-based and preference-based approaches at a population level. Cancer **120**(19), 2955–2964 (2014). https://doi.org/10.1002/cncr.28771

[18] Bertsimas, D., Dunn, J., Gibson, E., Orfanoudaki, A.: Optimal survival trees. Machine Learning (2022). https://doi.org/10.1007/s10994-021-06117-0

[19] Lee, K., Kruper, L., Dieli-Conwright, C.M., Mortimer, J.E.: The impact of obesity on breast cancer diagnosis and treatment. Current Oncology Reports **21**(5) (2019). https://doi.org/10.1007/s11912-019-0787-1

[20] Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S.: Random survival forests (2008). https://doi.org/10.48550/ARXIV.0811.1645

[21] Harrell, F.E.: Evaluating the yield of medical tests. JAMA: The Journal of the American Medical Association **247**(18), 2543 (1982). https://doi.org/10.1001/jama.1982.03320430047030

[22] Bamber, D.: The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of Mathematical Psychology **12**(4), 387–415 (1975). https://doi.org/10.1016/0022-2496(75)90001-2

[23] Parmigiani, G., Berry, D.A., Aguilar, O.: Determining carrier probabilities for breast cancer–susceptibility genes BRCA1 and BRCA2. The American Journal of Human Genetics **62**(1), 145–158 (1998). https://doi.org/10.1086/301670

[24] Gail, M.H., Greene, M.H.: Gail model and breast cancer. The Lancet **355**(9208), 1017 (2000). https://doi.org/10.1016/s0140-6736(05)74761-9

[25] Himes, D.O., Root, A.E., Gammon, A., Luthy, K.E.: Breast cancer risk assessment: Calculating lifetime risk using the tyrer-cuzick model. The Journal for Nurse Practitioners **12**(9), 581–592 (2016). https://doi.org/10.1016/j.nurpra.2016.07.027

[26] Ballard-Barbash, R., Taplin, S.H., Yankaskas, B.C., Ernster, V.L., Rosenberg, R.D., Carney, P.A., Barlow, W.E., Geller, B.M., Kerlikowske, K., Edwards, B.K., Lynch, C.F., Urban, N., Chrvala, C.A., Key, C.R., Poplack, S.P., Worden, J.K., Kessler, L.G.: Breast cancer surveillance consortium: a national mammography screening and outcomes database. American Journal of Roentgenology **169**(4), 1001–1008 (1997). https://doi.org/10.2214/ajr.169.4.9308451