DE NOVO CLASSIFICATION REQUEST FOR ALLIX5

REGULATORY INFORMATION

FDA identifies this generic type of device as:

Radiological software device to predict future breast cancer risk. A radiological software device to predict future breast cancer risk is a device that analyzes radiological images generated by breast imaging modalities and/or inputs derived from radiological images to provide qualified healthcare professionals with a prediction of the risk of future (incident) breast cancer. This device produces a numeric probability and/or risk category indicative of the patient's future breast cancer risk from the time the analyzed images were acquired. This device is not intended to diagnose, detect, or inform the treatment of cancer. The output of this device is not intended to guide interpretation of imaging exams.

NEW REGULATION NUMBER: 21 CFR 892.8500

CLASSIFICATION: Class II

PRODUCT CODE: SEZ

BACKGROUND

DEVICE NAME: Allix5

SUBMISSION NUMBER: DEN240047

DATE DE NOVO RECEIVED: September 6, 2024

SPONSOR INFORMATION:

Clairity, Inc.
201 W 5th St., Ste 1500
Austin, TX 78701

INDICATIONS FOR USE

The Allix5 is indicated as follows:

The Clairity Allix5 software device is intended to generate a 5-year risk prediction of breast cancer based on a bilateral screening mammogram. Allix5 provides a prediction of the percentage probability that the individual will receive a diagnosis of breast cancer or develop breast cancer within the 5-year timeframe following the screening mammogram, through analysis of mammography features and characteristics.

Eligible patients do not have a known breast cancer at presentation for their screening mammogram.

Allix5 is not intended to diagnose or detect breast cancer, or to provide care recommendations. Allix5 is not intended to replace or to be used as the sole determinant for clinical decision-making. Allix5 output is intended to be considered after the radiologist has completed the interpretation of the screening mammogram.

Allix5 analyzes full-field digital mammograms or directly acquired 2D images from Hologic Lorad Selenia and Selenia Dimensions Mammography Systems; it does not analyze synthetic-2D images.

LIMITATIONS

- The sale, distribution, and use of the Allix5 are restricted to prescription use in accordance with 21 CFR 801.109.
- Allix5 is not intended to diagnose or detect breast cancer, or to provide care recommendations.
- Allix5 is not intended to replace or to be used as the sole determinant for clinical decision-making.
- Allix5 can analyze only directly acquired 2-Dimensional screening mammography images from Hologic Lorad Selenia and Selenia Dimensions Mammography Systems.
- Allix5 scores greater than 5% should be interpreted with caution. While all scores greater than 5% are considered high risk, high risk scores produced by the model may tend to significantly overestimate risk, particularly scores above 10%.
- The effectiveness of Allix5 has not been demonstrated for the subpopulation of patients with breast implants.
- The device prediction of 5-year cumulative risk of breast cancer is unadjusted for competing risks of mortality.

PLEASE REFER TO THE LABELING FOR A COMPLETE LIST OF WARNINGS, PRECAUTIONS AND CONTRAINDICATIONS.

DEVICE DESCRIPTION

Allix5 is a Software as a Medical Device (SaMD) that takes as input two-dimensional screening mammography images acquired from Hologic Selenia Dimensions and Lorad Selenia systems and outputs a percentage probability that the patient will receive a diagnosis of breast cancer or develop breast cancer within the 5-year timeframe. The inputs required by the device are the four directly acquired 2D standard views (RCC, LCC, LMLO, RMLO) from the compatible mammography systems. Allix5 provides an output in the range of 0% - 100%. The Allix5 output is a population-based risk estimate reflecting the expected proportion of women, among those with similar AI-identified mammography features and characteristics, who will receive a diagnosis of or will develop a new breast cancer in the next 5 years. Allix5 outputs a percentage probability estimate of the pure risk of breast cancer based on a patient's screening mammogram,

where pure risk is the risk of breast cancer without adjustment for competing causes of mortality. Allix5's risk prediction algorithm was developed using a computer vision-based deep learning model that was trained on screening mammograms across diverse patient populations.

Allix5 is not intended to influence the interpretation of a screening mammogram in terms of detecting or diagnosing disease. Figure 1 illustrates how Allix5 is intended to be utilized within the current patient screening workflow:

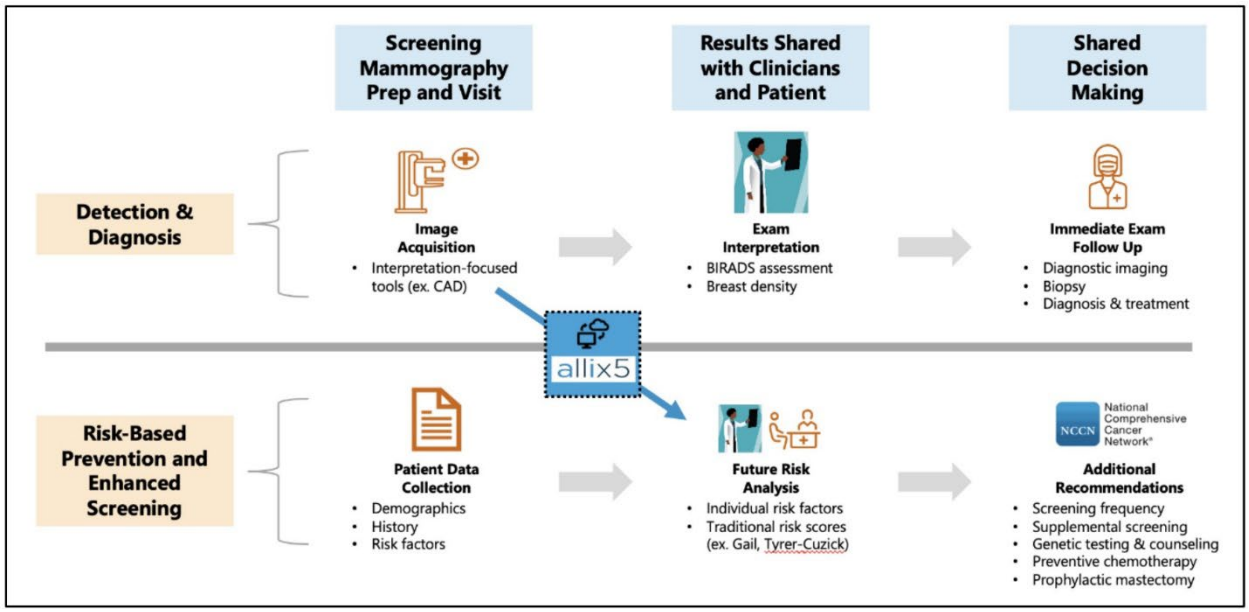

Figure 1. Clinical Workflow with Allix 5

The Allix5 device is designed to be integrated into existing clinical information systems that are already part of healthcare provider workflows; it does not have an independent user interface. The Allix5 device receives DICOM files from a PACS and automatically returns the Allix5 output via HL7. In this way, healthcare providers (HCPs) can access the Allix5 output alongside patient history and other relevant clinical data as they consider the patient's overall risk of breast cancer and make recommendations for screening and preventive care. HCPs can compare Allix5 outputs to established risk thresholds, such as those provided by the National Comprehensive Cancer Network (NCCN), when making care recommendations.

Allix5 can analyze only directly-acquired 2-Dimensional screening mammography images, and the compatible mammography systems are as follows:

- Selenia Dimensions from Hologic
- Lorad Selenia from Hologic

Algorithm Development

The Allix5 algorithm was trained on model development datasets collected from 22 clinical sites that are geographically diverse and represent a range of different types of screening centers (e.g., hospital-based and freestanding) with a diverse patient population. Table 1 below summarizes the distribution of the training datasets in the number of exams.

Clinical data records from each site were used to determine the ground truth. Ground truth for the cancer positive outcome within five years after the index screening mammogram is based on biopsy-proven breast cancer. Ground truth for cancer negative outcome at five years after the index screening mammogram is based on subsequent mammography and/or lack of evidence of a biopsy-proven breast cancer within the required time frame.

Table 1. Distribution of Training Dataset

Patient Demographics and Other Characteristics	Training Dataset (number of exams)
Age	
Age < 50	73,500
50 ≤ Age < 65	208,865
Age ≥ 65	83,551
Race	
White	225,309
Black	20,862
Asian	9,230
Indian	553
Hawaiian	440
Unknown	109,654
BI-RADS	
Positive	32,928
Negative	244,940
Unknown	8,8045
Breast Density	
Dense	170,398
Not Dense	160,974
Unknown	34,544

SUMMARY OF NONCLINICAL/BENCH STUDIES

SOFTWARE AND CYBERSECURITY

Allix5 software documentation and software verification and validation testing demonstrate that the device follows all recommendations for basic documentation level as outlined in the FDA guidance document, "Content of Premarket Submissions for Device Software Functions," issued on June 14, 2023. A description of the testing

protocols, including pass/fail criteria, and report of results was provided for the verification and validation activities and all testing results met design specifications and passed the acceptance criteria.

Allix5 cybersecurity documentation demonstrates that the device meets all the cybersecurity requirements as outlined in Section 524B of Federal Food, Drug, and Cosmetic Act (FD&C Act). This includes a threat model, software bill of materials, data security training, validation and mitigation of adversarial examples, cyber risk management, labeling, cyber testing, and post market cyber vulnerabilities and other information for safeguarding the algorithms.

SUMMARY OF CLINICAL INFORMATION

A multi-site retrospective powered study was conducted as a premarket pivotal study to demonstrate the predictive capability of the Allix5 device.

Clinical Sites and Patient Demographics

A total of 77,511 exams were collected from 44,112 patients in the following five validation sites and included in the pivotal study:

- Emory University Hospital Midtown in Atlanta, Georgia (hereafter Emory Midtown)
- Solis Mammography at Baylor Scott & White All Saints Medical Center in Fort Worth, Texas (hereafter Solis BASH)
- Solis Mammography Bedford at Lisa Trent Breast Center in Bedford, Texas (hereafter Solis Bedford)
- University of California, Davis Health Ellison Ambulatory Care Center in Sacramento, California (hereafter UCD Ellison)
- University of California, Davis Folsom Clinic in Folsom, California (hereafter UCD Folsom)

The selected testing sites are geographically diverse and represent a range of different types of screening centers and a population representing the U.S. screening mammography population. Use of these five clinical validation sites provided data that is independent of the data and sites used for algorithm training and tuning.

The study population included patients eligible for screening mammography, and the study exams included all four standard screening mammography views (RCC, LCC, LMLO, RMLO). Table 2 summarizes the data distribution for the pivotal study population.

Establishment of Ground Truth/Reference Standard

Ground truth for breast cancer occurrence within five years after the index screening mammogram was based on biopsy-proven breast cancer. Ground truth for breast cancer not occurring within five years after the index screening mammogram was based on subsequent mammography and/or lack of evidence of a biopsy-proven breast cancer within the required time frame.

Table 2. Data Distribution for the Pivotal Study Population

	Validation Study Exams		5-Year Cancer Positive Exams	
	Count	% of Total	Count	% of Total
Overall	77,511	100%	1,723	100%
Race				
Asian	5,735	7%	100	6%
Black or African American	28,931	37%	706	41%
Native American or Alaska Native	440	1%	9	1%
Native Hawaiian or Pacific Islander	385	0%	7	0%
White or Caucasian	34,743	45%	813	47%
Race Unknown	7,623	10%	96	6%
Ethnicity				
Hispanic	7,617	10%	128	7%
Non-Hispanic	55,986	72%	1,370	80%
Ethnicity Unknown	13,908	18%	225	13%
Age Group				
<50	20,156	26%	249	14%
≥50, <65	35,310	46%	725	42%
≥65	22,045	28%	749	43%
Breast Density				
Dense	30,912	40%	750	44%
Not Dense	44,627	58%	946	55%
Density Unknown	1,972	3%	27	2%
Breast Implants				
Present	1,738	2%	21	1%
Not Present	75,773	98%	1,702	99%
Mammography Model				
Lorad Selenia	22,093	29%	447	26%
Selenia Dimensions	55,491	72%	1,284	75%
Clinical Validation Site				
Emory Midtown	29,135	38%	730	42%
Solis Bash	11,400	15%	198	11%
Solis Bedford	9,682	12%	224	13%
UCD Ellison	21,195	27%	451	26%
UCD Folsom	6,099	8%	120	7%

The precise time of breast cancer onset is unknown. In establishing ground truth, two types of censoring of the missing time to breast cancer onset were taken into consideration:

- For exams that are followed by a biopsy-proven breast cancer within five years, the time of breast cancer onset is likely to have occurred in the time period between the subject's last known negative exam and the positive biopsy.

- For exams without a biopsy-proven breast cancer within 5 years nor a negative exam at or after 5 years, the potential existence of breast cancer and, if it exists, whether the time of its onset is before or after 5 years since the index screening mammogram, is uncertain.

To account for time to breast cancer onset being censored, the time interval between a negative exam and a cancer diagnosis was used to estimate each index screening mammogram's 5-year cumulative probability of breast cancer occurrence since the index screening mammogram. This probability constitutes "ground truth", replacing the 0/1 (positive/negative) labels used in estimating sensitivity and specificity without censoring. The probabilities of being breast cancer positive by t = 5 years since the index screening mammogram may be categorized as either "positive," "negative," or "ambiguous.":

- **Positive Case**. The probability of being cancer positive within t = 5 years since the index screening mammogram is 1 when a biopsy-confirmed breast cancer diagnosis occurred within 5 years of the index screening mammogram.
- **Negative Case**. The probability of being cancer positive within t = 5 years since the index screening mammogram is 0 when either
    - Breast cancer was never biopsy-confirmed, and the subject had a screening mammogram, diagnostic mammogram, or negative breast biopsy 5 years or more after the index exam, or
    - Breast cancer was biopsy-confirmed more than 5 years after the index screening mammogram and the subject had a screening or diagnostic mammogram prior to the cancer diagnosis that occurred 5 years or more after the index screening mammogram which did not result in biopsy-confirmed breast cancer.
- **Ambiguous Case**. The probability of being cancer positive within t = 5 years since the index screening mammogram is between 0 and 1 when either
    - Breast cancer was never biopsy-confirmed, and the subject did not have a screening mammogram, diagnostic mammogram, or negative breast biopsy 5 years or more after the index exam, or
    - Breast cancer was biopsy-confirmed more than 5 years after the index exam, but the subject did not have an earlier screening or diagnostic mammogram that occurred 5 years or more after the index screening mammogram which did not result in biopsy-confirmed breast cancer.

Study Objectives

The primary objective of the study was to evaluate the discrimination and calibration of the Allix5 device's prediction of breast cancer risk within the 5-year timeframe following a screening mammography exam in the total study population, which represents the intended use population of the general screening population. In addition to the performance in the total study population, performance was also evaluated on the total study population excluding patients with

current cancer. Current cancers were defined as a biopsy-proven breast cancer diagnosed in the 90 days[1] following a positive BI-RADS assessment (i.e., BI-RADS score of 0, 3, 4, or 5) on the index screening mammogram.

The Allix5's performance in the total study population excluding patients with current cancer was evaluated for several reasons. First, this group of patients represents the vast majority of the screening population. Second, the probable benefits of the device but also its probable risks (i.e., leading to possible harm) due to device misprediction of cancer risk are considered greater in this group than the group of patients with current cancer. Third, the risk score for patients without current cancer informs supplemental screening and/or any other risk reduction measures. Regarding probable risks, underestimating the breast cancer risk for patients without current cancer may preclude the patient from supplemental screening, which may further delay the diagnosis of breast cancer that would have otherwise been detected earlier. Conversely, overestimating breast cancer risk for patients without current cancer may cause unnecessary anxiety, and lead to prescription of unnecessary supplemental screening exams and/or chemopreventive medication. Such risks may not apply to patients with current cancer, as these patients are managed differently from patients without current cancer in the clinical workflow, and their clinical management may not be influenced as much by the device risk score as would patients without current cancer.

Co-primary Endpoints

- Clinically acceptable discrimination of Allix5 in the total study population as measured by the area under the time-dependent, cumulative-dynamic receiver operating characteristic (ROC) curve (AUC);
- Clinically acceptable calibration of Allix5 in the total study population as evaluated by the Greenwood-Nam-D'Agostino (GND) test, modified to test if the miscalibration of device risk prediction is (in an average sense) within accepted tolerance bounds for miscalibration.

Statistical Methods

To evaluate the discrimination co-primary endpoint, the lower limit of the 95% confidence interval on Allix5's area under the time-dependent, cumulative-dynamic receiver operating characteristic (ROC) curve, denoted AUC(t) for follow-up time t and evaluated at t=5 years since the index screening mammogram, was compared with a performance goal of 0.64. The performance goal of 0.64 was selected based on the discrimination performance of traditional risk models [1-3]. Statistical methods were used to adjust the estimation of AUC(5) for interval censoring of the precise time of breast cancer onset [4-5].

To assess the calibration co-primary endpoint, the GND test statistic was used to evaluate whether the predicted event rate, that is, the device model prediction of the 5-year, cumulative risk (probability) of breast cancer, was adequately calibrated with the observed event rate. To make this assessment, exams were divided into ten groups representing deciles of device model

[1] Note that a 90-day window for positive biopsy findings was used in attempting to exclude patients who are not diagnosed with cancer as a result of further workup prompted by findings on the index screening mammogram [6-9].

predictions. The mean model prediction in each decile was compared with the Turnbull estimate of the event rate in the decile, which adjusts for left-, right-, and interval censoring of the unknown, precise time of breast cancer onset [10]. Per decile, the squared difference between the mean prediction and the Turnbull estimate of the event rate was standardized by dividing by the bootstrap variance estimate of the Turnbull estimate. The Turnbull estimate and its variance were used in place of the commonly used Kaplan-Meier variance estimate of the event rate and its Greenwood variance estimate because they are designed for right-censoring, but not also for left- and interval-censoring of event times occurring in the dataset [11]. The GND test statistic is the sum of these standardized statistics across the 10 deciles. Under perfect calibration, the GND test statistic follows the chi-square distribution with 10 degrees of freedom (dof). In the modified version of the GND test, lower and upper bounds on the expected difference between the mean prediction and the estimated event rate were pre-specified per decile to represent clinically acceptable tolerances for miscalibration in the intended use population. The lower and upper bounds on the difference were symmetric, i.e., upper bound = – lower bound. Under the null hypothesis that in each decile the expected difference between predicted and estimated event rates is equal to the lower or upper bound, the GND statistic follows a non-central chi-square distribution with 10 dof, with the non-centrality parameter determined from the bounds and the data. If the test statistic is less than the 2.5^{th} percentile of this non-central chi-squared distribution ($p < 0.025$), the null hypothesis is rejected and agreement of mean prediction with observed risk is declared significantly within the calibration bounds (in an average sense) across the deciles.

Secondary Endpoints

The same study endpoints, statistical methods and performance goals used for the total study population were repeated for a subset of the study population excluding patients with current cancer, and for subsets of White patients and Black patients presented for screening mammography.

The secondary endpoint for testing on the total study population excluding patients with current cancer was selected due to the reasons noted above. The secondary endpoint for testing on subgroups of White patients and Black patients was selected because literature shows that traditional mathematical models typically have relatively lower performance in Black patients, as those models were developed using data from a relatively homogeneous, predominantly White population [1, 12-13]. As a result, there was a need to evaluate Allix5 performance on these subgroups to see if it provides greater benefit to these subpopulations, as compared with the performance of the traditional mathematical models.

Subgroup Analyses

The Allix5 discrimination and calibration performance was also evaluated in subgroups defined by race/ethnicity (Asian, Hispanic, Black, White), age group (< 50, ≥50 to <65, ≥ 65), breast implant status (present or not), mammography systems and models (Hologic Lorad Selenia, Hologic Selenia Dimensions), and clinical validation site (Emory Midtown, Solis Bash, Solis Beford, UCD Ellison, UCD Folsom).

Pivotal Study Results

Table 3 shows the pivotal study results using performance metrics noted above in the total study population (co-primary endpoints), in the total study population excluding current cancers (secondary endpoints) and in the subgroups defined above for the total study population.

Table 3. Discrimination and Calibration Results for the Co-primary and Secondary Endpoints for the Total Study Population and for Subgroup Analyses

		Discrimination:	Calibration:		
Study Population	**Exams**	**AUC(5) [95% CIs]**	**GND Test Statistic**	**Non-centrality parameter**	***P* value**
Formally-Tested Endpoints					
Total Study Population	77,511	0.70 [0.69, 0.72]	21.70	121.25	< .0001
Excluding Current Cancers	77,072	0.66 [0.66, 0.68]	95.27	132.08	0.0174
Sub-group Analyses (Total Study Population)					
Race & Ethnicity					
Asian	5,735	0.72 [0.67, 0.77]	13.45	11.80	0.1830
Black	28,931	0.67 [0.65, 0.69]	24.81	43.29	0.0107
White	34,743	0.73 [0.71, 0.74]	26.92	56.06	0.0016
Hispanic	7,617	0.70 [0.64, 0.75]	7.99	14.37	0.0139
Age Group					
<50	20,156	0.68 [0.66, 0.72]	24.66	34.43	0.0498
≥50, <65	35,310	0.69 [0.67, 0.71]	9.33	54.35	0.0000
≥65	22,045	0.68 [0.66, 0.70]	46.80	38.11	0.5202
Breast Density					
Dense	30,912	0.69 [0.67, 0.71]	26.17	42.88	0.0168
Not Dense	44,627	0.71 [0.69, 0.73]	20.99	83.14	< .0001
Breast Implants					
Present	1,738	0.73 [0.59, 0.87]	5.46	2.08	0.1153
Not Present	75,773	0.70 [0.69, 0.72]	22.30	119.48	< .0001
Mammography Model					
Lorad Selenia	22,093	0.72 [0.70, 0.75]	9.89	42.04	< .0001
Selenia Dimensions	55,491	0.70 [0.68, 0.72]	29.01	77.52	< .0001

Based on the results shown in Table 3, the hypothesis testing conducted in the pivotal study demonstrates that, for the total study population, Allix5 met the discrimination co-primary endpoint as supported by the reported AUC(5) of 0.70 with the 95% confidence interval (0.69,

0.72) being above the performance goal of 0.64, and met the calibration co-primary endpoint, as supported by modified GND p-value < 0.001. For the secondary endpoints defined for the total study population excluding patients with current cancer, Allix5 also met the discrimination endpoint, as supported by the report AUC(5) of 0.66 with the 95% confidence interval (0.66, 0.68) being above 0.64, and met the calibration endpoint, as supported by p-value = 0.0174.

Calibration plots with modified GND statistics are presented in Figure 2 for the total study population and the total study population excluding current cancers, respectively. While the calibration plot shown in Figure 2 (right) shows acceptable calibration for the subset of patients excluding current cancers for device predicted risk scores up to 5%, it also indicates significant overestimation in the last decile of highest predicted risk. Clinically, overestimating a truly high risk as an even higher falsely elevated risk may not have much impact on patient management in that clinical practice guidelines such as the current National Cancer Comprehensive Network (NCCN) guidelines [14-15] typically recommend increased screening or other risk prevention measures when the 5-year risk of breast cancer exceeds a risk threshold considerably lower than 5%, e.g., 1.7% or 3.0%. Nonetheless, to mitigate the risk of patient mismanagement due to overestimation of high risk, a cautionary statement is included in the device labeling and device output informing the end user how to interpret risk scores above 5%.

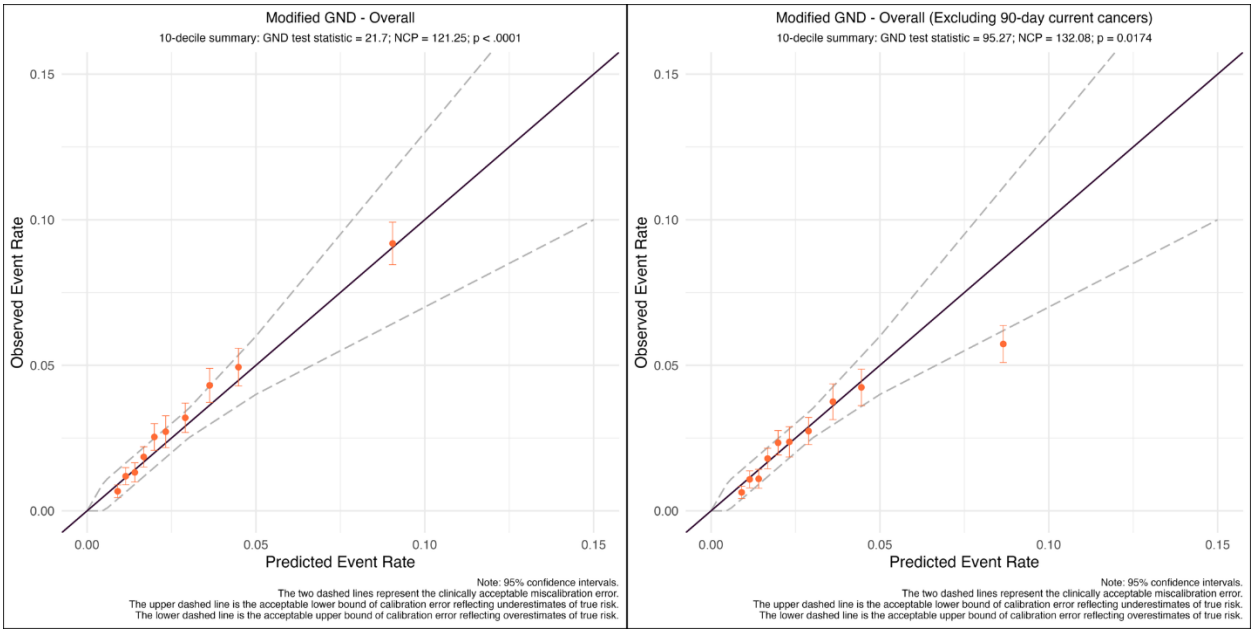

Figure 2. Calibration scatterplots of observed event rate (Turnbull estimate) vs. predicted event rate (mean model prediction) per decile of model prediction for the total study population (left) and the total study population excluding patients with current cancer (right). The identity line (solid) and the tolerance bounds for acceptable miscalibration (dashed) are also depicted on the plots. The heading of each plot shows the modified GND test results for whether the model predictions may be declared significantly within the calibration bounds.

Table 3 shows the study results for the subgroups defined above in the Total Study Population, whereas Table 4 shows the study results for these subgroups excluding patients with current cancer.

Based on the results shown in Table 3, for the subpopulations of patients who have breast implants, are Asian, are less than 50 years of age, or are greater than 65 years of age, the calibration of device model predictions with observed risk was not significantly within the bounds for acceptable calibration (modified GND p value > 0.025).

In particular, as shown in Table 3, for the subpopulation of patients with breast implants, the effectiveness of Allix5 was not demonstrated. The AUC(5) estimate of 0.73 was not significantly greater than the acceptance criterion of 0.64 because the 95% confidence interval of (0.59, 0.87), contains 0.64 due to large uncertainty in the estimate. Also, calibration of device model prediction with observed risk was not significantly within the calibration bounds because the GND test p value of 0.1153 is > 0.025. Similar trends were observed in Table 4 for the subpopulation of patients who have breast implants and do not have current cancer.

Table 4. Subgroup Analysis for the Total Study Population Excluding Patients with Current Cancer

		Discrimination:	Calibration:		
Study Population	**Exams**	**AUC(5) [95% CIs]**	**GND Test Statistic**	**Non-centrality parameter**	***P* value**
Sub-group Analyses (Excluding Current Cancers)					
Race & Ethnicity					
Asian	5,708	0.69 [0.63, 0.77]	23.27	13.40	0.5834
Black	28,761	0.63 [0.59, 0.66]	61.04	49.58	0.5915
White	34,521	0.68 [0.66, 0.70]	41.64	62.84	0.0209
Hispanic	7,589	0.67 [0.61, 0.73]	23.46	16.21	0.4636
Age Group					
<50	20,091	0.64 [0.60, 0.67]	73.20	38.84	0.9622
≥50, <65	35,124	0.65 [0.63, 0.68]	56.09	60.99	0.1991
≥65	21,857	0.64 [0.61, 0.66]	69.98	50.61	0.7680
Breast Density					
Dense	30,720	0.65 [0.62, 0.67]	31.47	43.83	0.0475
Not Dense	44,383	0.67 [0.65, 0.69]	79.19	90.53	0.1466
Breast Implants					
Present	1,733	0.63 [0.49, 0.76]	8.88	2.50	0.3508
Not Present	75,339	0.66 [0.65, 0.68]	105.34	135.85	0.0394
Mammography Model					
Lorad Selenia	21,975	0.68 [0.65, 0.71]	55.57	48.34	0.4781
Selenia Dimensions	55,164	0.66 [0.64, 0.68]	47.24	82.14	0.0035

The calibration plots for patients who have breast implants shown in Figure 3 provide a visual assessment of the deviation of per decile mean model prediction from observed risk and the uncertainty of the deviation. The large uncertainties are due to a small sample size and a small number of breast cancers in this subpopulation. To mitigate the risk of patient mismanagement due to inaccurate calibration of risks for patients with breast implants, a cautionary statement was included in the device labeling and device outputs informing the user that the effectiveness of Allix5 has not been demonstrated for the subpopulation of patients with breast implants.

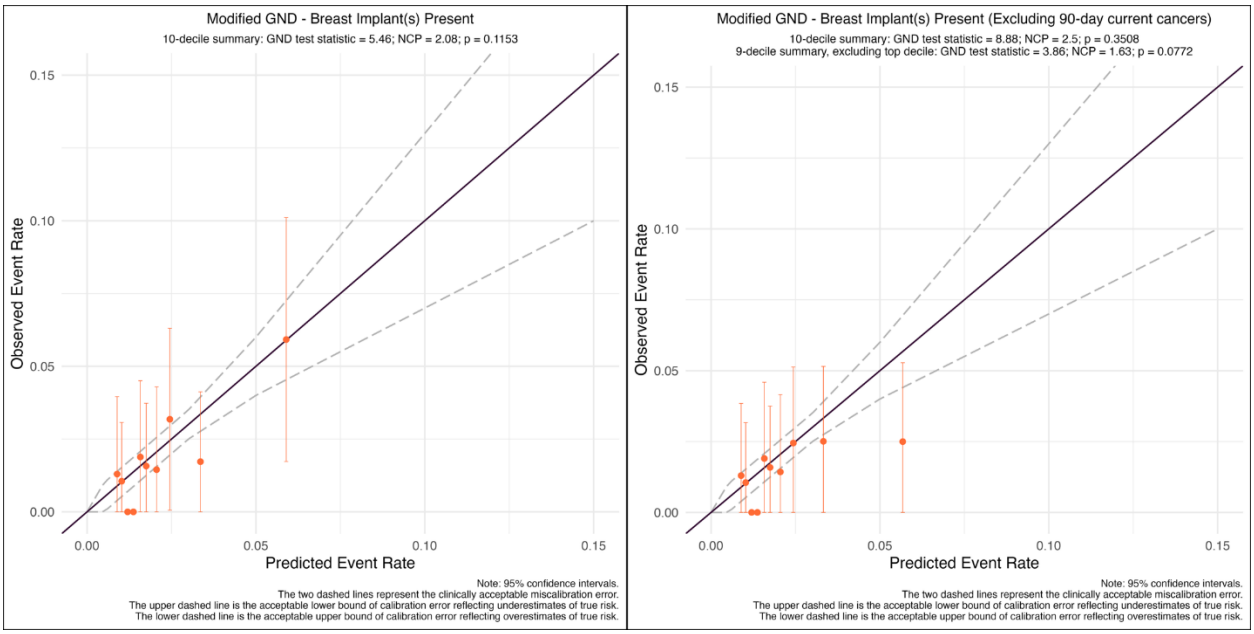

Figure 3. Calibration Plots for the Subpopulation with Breast Implants (left) and for the Subpopulation with Breast Implants Excluding Patients with Current Cancers (right)

Based on the results shown in Table 3 and Table 4, acceptable discrimination or calibration of Allix5 was not demonstrated in the subpopulations noted above for reasons that may have included increased chances for falsely insignificant results with multiple testing, insufficient subgroup data (small sample size and/or number of breast cancers), and/or lack of device effectiveness.

Table 5 describes the mean model prediction, observed Turnbull estimate, calibration error and 95% confidence interval of calibration error for each decile for the total study population, and Table 6 shows these decile-level statistics for the total study population excluding current cancers.

In summary, based on the results presented above, Allix5 met (1) the discrimination co-primary endpoint and the calibration co-primary endpoint for the total study population, (2) the secondary endpoint defined by discrimination and calibration for the total study population excluding patients with current cancer, and (3) the secondary endpoint defined by discrimination and calibration for both the White and Black populations in the total study population. Based on the subgroup analysis results, clinical acceptable discrimination was demonstrated for all the subpopulations of patients except for the subpopulation of patients who have breast implants;

clinical acceptable calibration was demonstrated for the other subpopulations of patients but not demonstrated for the subpopulations of patients who have breast implants, are Asian, are less than 50 years of age, or are greater than 65 years of age.

Table 5. Decile-level Statistics for the Total Study Population

Decile	Mean model prediction	Observed Turnbull estimate	Calibration error (95% CI)
1	0.91%	0.67%	0.23% (0.01%, 0.45%)
2	1.14%	1.19%	-0.05% (-0.34%, 0.24%)
3	1.41%	1.32%	0.09% (-0.24%, 0.42%)
4	1.68%	1.85%	-0.17% (-0.52%, 0.18%)
5	1.99%	2.54%	-0.55% (-1.00%, -0.10%)
6	2.33%	2.72%	-0.39% (-0.94%, 0.16%)
7	2.91%	3.20%	-0.29% (-0.80%, 0.22%)
8	3.63%	4.31%	-0.68% (-1.27%, -0.09%)
9	4.48%	4.94%	-0.45% (-1.10%, 0.20%)
10	9.05%	9.19%	-0.14% (-0.87%, 0.59%)

Table 6. Decile-level Statistics for the Total Study Population Excluding Patients with Current Cancer

Decile	Mean model prediction	Observed Turnbull estimate	Calibration error (95% CI)
1	0.91%	0.64%	0.27% (0.05%, 0.49%)
2	1.14%	1.08%	0.06% (-0.23%, 0.35%)
3	1.41%	1.10%	0.31% (-0.02%, 0.64%)
4	1.68%	1.80%	-0.12% (-0.47%, 0.23%)
5	1.99%	2.34%	-0.35% (-0.76%, 0.06%)
6	2.32%	2.37%	-0.05% (-0.56%, 0.46%)
7	2.89%	2.74%	0.15% (-0.32%, 0.62%)
8	3.61%	3.75%	-0.14% (-0.75%, 0.47%)
9	4.45%	4.24%	0.21% (-0.42%, 0.84%)
10	8.65%	5.73%	2.92% (2.29%, 3.55%)

Reproducibility Testing

Reproducibility testing was conducted to demonstrate the robustness of Allix5's risk scores to (1) variabilities associated with repeated screening mammography acquisitions and (2) alternative image selection within a single exam.

The reproducibility study #1 used a dataset of 1,046 patients with longitudinal mammograms across three U.S. institutions (Colorado, Georgia, and Washington, D.C.). The patient images were transformed in various ways to be representative of variations observed in real patient data encountered in clinical mammography practice. Specifically, affine and photometric image

transformations were computed based on longitudinal exams and applied to a separate test cohort of 102 patients. A random selection of 10 transforms were applied to each of the 102 test exams, yielding 1020 transformed exams for analysis. Results demonstrated that the reliability of Allix5 under simulated repeated screening exams was within acceptable reproducibility standards for imaging biomarkers, supported by a mean Intraclass Correlation Coefficient (ICC) of 98%, a mean absolute risk difference of 0.008 (median 0.002), a mean Coefficient of Variation of 17.5%, and the mean Standard Error of Measurement of 0.0024.

Reproducibility study #2 assessed whether Allix5's 5-year breast cancer risk predictions remain consistent when the selected mammographic image for a given standard view is replaced by another valid image from the same exam. The study used 2400 total exams, which consists of 600 images for each of the four standard views (LCC, LMLO, RCC, RMLO) sampled across four U.S. regions. During the evaluation, each view was tested by substituting its selected image with a randomly chosen alternative valid image from the same exam. Acceptable reliability was demonstrated by ICC values ranging from 0.83 to 0.92 (above the acceptance threshold of 0.75), and acceptable stability was demonstrated by a mean absolute difference of 0.0043.

Pediatric Extrapolation

In this De Novo request, existing clinical data were not leveraged to support the use of the device in a pediatric patient population. The device requires a screening mammogram as an input, and clinical practice guidelines do not recommend routine screening mammography for pediatric patients.

POST-MARKET MONITORING

A post-market device monitoring plan was provided to ensure that changes to the hardware and/or software of the compatible mammography systems will be monitored and any changes that could impact Allix5's performance will be identified. Additionally, to ensure Allix5's consistent performance across diverse clinical environments, the post-market device monitoring plan was designed to monitor site-specific statistics that may be potential causes of changes in performance or outputs indicating drift. Data will be collected through ongoing, active systematic collection, and monitoring will include analysis and interpretation of data.

LABELING

The labeling meets the requirements of 21 CFR 801.109 for prescription devices and includes information on device inputs and output, instructions for use, intended patient population and intended users of the device, as well as adequate warnings and precautions. In particular, as mentioned in the Pivotal Study Results, cautions related to the effectiveness of Allix5 on the subpopulation of patients with breast implants and how to interpret Allix5 scores greater than 5% were included to mitigate the risk of misinterpretation. A summary of the performance testing and a discussion of the performance data including subgroup analyses was also provided in the user manual to facilitate the understanding of the device benefits, limitations, and risks to the end user.

RISKS TO HEALTH

The table below identifies the risks to health that may be associated with use of a radiological software device to predict future breast cancer risk and the measures necessary to mitigate these risks.

Risks to Health	Mitigation Measures
Device provides inaccurate risk discrimination and/or calibration contributing to: • Falsely high risk output, leading to patients receiving unnecessary additional imaging or preventive risk reduction measures • Falsely low risk output, leading to patients not receiving additional imaging or preventive risk reduction measures	Clinical performance testing Software verification, validation, and hazard analysis Labeling Postmarket monitoring plan
Inconsistent device output due to differences in acquisition, equipment settings, or other factors affecting image selection or quality	Clinical performance testing Software verification, validation and hazard analysis Labeling Postmarket monitoring plan
User misinterpretation of device result(s) contributing to: • Inappropriate categorization of patient risk resulting in erroneous patient management decisions related to additional imaging or preventive risk reduction measures • Overreliance on device output Patient misinterpretation of device result(s) contributing to • Considering low risk to be negligible which may contribute to not receiving standard-of-care screening or preventive risk reduction measures	Labeling Postmarket monitoring plan

SPECIAL CONTROLS

In combination with the general controls of the FD&C Act, the radiological software device to predict future breast cancer risk is subject to the following special controls:

(1) Data obtained from clinical performance validation testing acquired under anticipated conditions of use must demonstrate that the device performs as intended when used in the intended patient population. Documentation must include the following:

(i) A description of prespecified performance testing protocols (including the study objectives, study endpoints, statistical hypotheses, performance goals, sample size calculation, and statistical analyses, including adjustment for left- and/or right-censoring of the time of disease onset). Performance goals used to determine success of the clinical validation study must be clinically justified;
(ii) A description of the dataset(s) used. Validation must use a clinical test dataset acquired from a representative patient population. The test dataset must be representative of the range of data sources and data quality likely to be encountered in the intended use population and relevant use conditions in the intended use environment. The test dataset must be independent of the data used in the training/development of the device;
(iii) Establishment of a reference standard, with clinical justification, to distinguish those who experience the disease by the designated future time from those who do not, or a time interval within which the disease onset is known to have occurred or not occurred. Study protocols must include a description of the methodology for determining the reference standard for training and test datasets;
(iv) Results to validate the device output pertaining to future disease risk prediction. The performance assessment must be based on clinically justified measures of discrimination and calibration of risk categorization or score estimates in the intended patient population. Agreement between device predicted risk outputs and observed risk must be calculated across the range of risks expected in the intended patient population;
(v) The clinical performance of discrimination and calibration of device outputs must demonstrate the generalizability of device performance across clinically important subgroups. Subgroup analysis of discrimination and calibration by study site, relevant demographic subgroups, image acquisition system, and any other applicable confounders of clinical interest must be provided; and
(vi) Data must demonstrate reproducibility of the device output across the range of input image acquisition settings including acquisition equipment and patient positioning.

(2) Software verification, validation, and hazard analysis must be provided. Software documentation must include:
(i) A technical description of the model/algorithm(s) and algorithm inputs and outputs; and
(ii) Verification and validation data that demonstrate software ensures input radiological images are adequate for processing.

(3) Labeling must include:
(i) Compatible imaging requirements for input;
(ii) A warning that the output of this device is not intended to guide interpretation of imaging exams;
(iii) A warning that the user of this device should consider other clinical information for patient management;

(iv) A warning that the device is not intended to diagnose, detect, or inform the treatment of disease;
(v) A summary of the clinical performance testing methods, including results of the performance testing for tested performance measures/metrics, selection criteria, truthing, patient dataset characteristics, and subgroup analyses by relevant confounders;
(vi) A description of output reproducibility and results of reproducibility testing;
(vii) Device limitations or a description of subpopulations for which the device may not perform as expected or for whom the device has not been validated; and
(viii) A summary of the device's current performance that incorporates clinical performance testing and data collected from post-market performance monitoring.

(4) The device manufacturer must develop and implement a post-market performance management plan that ensures regular assessment of the generalizability and device performance in the intended patient population in real-world use. The plan must include:
(i) Data collection, analysis methods, and procedures for:
(A) Monitoring relevant performance characteristics and detecting changes in performance;
(B) Identifying sources of performance changes between validation and real-world environment over time; and
(C) Assessing the results from the performance testing on safety and effectiveness;
(ii) Procedures for communicating the device's current performance to the users.

BENEFIT-RISK DETERMINATION

The risks of the device are based on nonclinical study information as well as data collected in a clinical study described above. The probable risks of the device are: (1) falsely high or low risk outputs by the device, (2) inconsistent device output, and (3) user's and/or patient's misinterpretation of device result(s). If a patient is put into a higher risk category than they should have been in due to inaccurate, inconsistent device outputs, and/or user misinterpretation, this could contribute to patients receiving unnecessary supplemental screening and/or preventive risk reduction measures. If a patient is put into a lower risk category than they should have been in, this could contribute to patients not receiving supplemental screening or preventive risk reduction measures. If a patient misinterprets the device result(s) and considers low risk to be negligible, this may contribute to the patient not receiving standard-of-care screening, additional imaging, or preventive risk reduction measures.

The probable benefits of the device are based on the nonclinical study as well as data collected in a clinical study as described above. The probable benefits of the device include (1) clinically acceptable risk discrimination due to Allix5 using image-based algorithms that could eliminate errors from inaccurately patient-reported or missing input used by traditional mathematical models, (2) clinically acceptable risk calibration due to device development based on data from

a population representing the U.S. screening mammography population, and (3) clinically acceptable discrimination and calibration performance on patients with dense breast tissue.

The current standard of care uses traditional mathematical risk models such as the Gail Model and Tyrer-Cuzick Model. These models rely on self-reported patient data, which can have missing or inaccurate data points that may skew the results. In contrast, Allix5 is image-based, which eliminates errors from inaccurately patient-reported or missing inputs used in the conventional mathematical models. This benefit is supported by the superior discrimination result (AUC(5): 0.70) with 95% confidence interval (0.69, 0.72) being above the performance goal of 0.64, informed by the traditional risk models. Traditional mathematical models also vary considerably in their risk categorization, and there is evidence that they tend to have racial and ethnic bias, as they were developed using data from a relatively homogeneous, predominantly white population [1, 12-13]. Based on literature, traditional mathematical models showed less accurate performance in Black, Asian and Hispanic subpopulations [2-3, 16]. In contrast, Allix5's algorithm was trained and developed using data from a population which more accurately represents the U.S. screening mammography population than that used in the traditional risk models. This benefit is supported by the clinically acceptable calibration performance on (1) the total study population, (2) the total study population excluding patients with current cancers, and (3) the Black and Hispanic subpopulations, as shown in Table 3. Among patients with dense breast tissue (a very common tissue type which is a risk factor for breast cancer development), the healthcare provider can use the Allix5 output in distinguishing between patients who have sufficiently high risk to warrant supplemental screening and/or other interventions, and patients who do not have sufficiently high risk. This is a benefit over the traditional mathematical models since most of the traditional mathematical models do not use breast density to calculate the predicted risk [12]. The Allix5's benefit on patients with dense breast tissue is supported by the clinically acceptable discrimination and calibration performance in the subpopulation of patients with dense breast tissue, as shown in Table 3.

To characterize the uncertainty in the device performance, discrimination and calibration were evaluated across important subgroups, as shown in Table 3. The device demonstrated clinically acceptable risk discrimination and risk calibration for the following subpopulations: (1) total study population excluding patients with current cancer, (2) White/Black/Hispanic patients, (3) patients with dense/non-dense breast tissue, (4) patients who do not have breast implants, (5) patients who are greater than 50 years and less than 65 years of age. However, acceptable calibration was not demonstrated for the subpopulations of patients who have breast implants, are Asian, are less than 50 years of age, or are greater than 65 years of age. Additionally, the device risk scores for the subset of patients excluding patients with current cancer significantly overestimated in the last decile of highest predicted risk. As noted in the discussion of performance data, clinically, overestimating a truly high risk as an even higher falsely elevated risk may not have much impact on patient management. That said, to further mitigate those risks, warnings and cautionary statements were included in the device labeling and device output to inform the user of the device limitations. A detailed summary of the device performance data was also included in the device labeling to inform the user of the lack of device effectiveness in these subgroups. Additionally, to ensure consistent performance across diverse clinical environments (including changes in settings of imaging equipment), a post-market device

monitoring plan was developed to monitor Allix5's performance in the clinical sites and detect data drift. Given these mitigations, the uncertainty of Allix5's probable risks is further reduced.

Patient Perspectives

This submission did not include specific information on patient perspectives for this device.

Benefit/Risk Conclusion

In conclusion, given the available information above, for the following indication statement:

> The Clairity Allix5 software device is intended to generate a 5-year risk prediction of breast cancer based on a bilateral screening mammogram. Allix5 provides a prediction of the percentage probability that the individual will receive a diagnosis of breast cancer or develop breast cancer within the 5-year timeframe following the screening mammogram, through analysis of mammography features and characteristics.
>
> Eligible patients do not have a known breast cancer at presentation for their screening mammogram.
>
> Allix5 is not intended to diagnose or detect breast cancer, or to provide care recommendations. Allix5 is not intended to replace or to be used as the sole determinant for clinical decision-making. Allix5 output is intended to be considered after the radiologist has completed the interpretation of the screening mammogram.
>
> Allix5 analyzes full-field digital mammograms or directly acquired 2D images from Hologic Lorad Selenia and Selenia Dimensions Mammography Systems; it does not analyze synthetic-2D images.

The probable benefits outweigh the probable risks for the Allix5. The device provides benefits, and the risks can be mitigated by the use of general controls and the identified special controls.

CONCLUSION

The De Novo request for the Allix5 is granted and the device is classified as follows:

> Product Code: SEZ
> Device Type: Radiological software device to predict future breast cancer risk
> Regulation Number: 21 CFR 892.8500
> Class: II

REFERENCES

1. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004 Apr 15;23(7):1111-30. doi: 10.1002/sim.1668. Erratum in: Stat Med. 2005 Jan 15;24(1):156. PMID: 15057881.

2. Rosner BA, Colditz GA, Hankinson SE, Sullivan-Halley J, Lacey JV Jr, Bernstein L. Validation of Rosner-Colditz breast cancer incidence model using an independent data set, the California Teachers Study. Breast Cancer Res Treat. 2013 Nov;142(1):187-202. doi: 10.1007/s10549-013-2719-3. Epub 2013 Oct 26. PMID: 24158759; PMCID: PMC3825503.
3. McCarthy AM, Guan Z, Welch M, Griffin ME, Sippo DA, Deng Z, Coopey SB, Acar A, Semine A, Parmigiani G, Braun D, Hughes KS. Performance of Breast Cancer Risk-Assessment Models in a Large Mammography Cohort. J Natl Cancer Inst. 2020; 112(5):489–497.
4. Martínez-Camblor P, Bayón GF, Pérez-Fernández S. Cumulative/dynamic ROC curve estimation, Journal of Statistical Computation and Simulation, 86:17, 3582-3594, doi: 10.1080/00949655.2016.1175442.
5. Díaz-Coto S, Martínez-Camblor P, Corral-Blanco NO (2020). Cumulative/dynamic ROC curve estimation under interval censorship, Journal of Statistical Computation and Simulation, 90:9, 1570-1590, doi: 10.1080/00949655.2020.1736071.
6. Lawson BA, Bissell MCS, Miglioretti, DL. Multilevel Factors Associated With Time to Biopsy After Abnormal Screening Mammography Results by Race and Ethnicity. JAMA Oncol 2022, 8(8):1115-1126. doi:10.1001/jamaoncol.2022.1990.
7. Vijayaraghavan GR, Guembou IM, Vedantham S. The Current State of Timeliness in the Breast Cancer Diagnosis Journey: Abnormal Screening to Biopsy. Seminars in Ultrasound, CT and MRI (2023) 44(1):56-61.
8. Donovan CA, Kaufman CS, Thomas KA, Polat AK, Thomas M, Mack B, Gilbert A, Sarantou, T. Timeliness of Breast Diagnostic Imaging and Biopsy in Practice: 15 Years of Collecting, Comparing, and Defining Quality Breast Cancer Care. Ann Surg Oncol (2023) 30:6070-6078.
9. Oluyemi ET, Grimm LJ, Goldman L, Simanowith M, Yao K, Rosenberg RD. Rate and Timeliness of Diagnostic Evaluation and Biopsy After Recall From Screening Mammography in the National Mammography Database. Clinical Practice Management (2024). 21 (3): 427-438.
10. Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. J. R. Stat. Soc. Ser. B, 38(3), 290–295.
11. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. Stat Med. 2015 May 10;34(10):1659-80. doi: 10.1002/sim.6428. Epub 2015 Feb 11.
12. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81(24):1879-1886. doi:10.1093/jnci/81.24.1879
13. Cuzick J, Forbes J, Edwards R, et al. First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial. Lancet. 2002;360(9336):817-824. doi:10.1016/s0140-6736(02)09962-2
14. NCCN Guidelines. (2022, January 31). Breast Cancer Risk Reduction. Retrieved from NCCN.org: https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1420.
15. NCCN Guidelines. (2021, May 6). Breast Cancer Screening and Diagnosis. Retrieved from NCCN.org: https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1421.

16. Gail MH. Choosing Breast Cancer Risk Models: Importance of Independent Validation. J Natl Cancer Inst. 2020 May 1;112(5):433-435. doi: 10.1093/jnci/djz180. PMID: 31556449; PMCID: PMC7225665.