WHITE PAPER

# Estimating the Likelihood of Near-Term Breast Cancer using ProFound AI® Risk

## Theory and Clinical Performance



DMM279 Rev 1

## Authors

**Jim Pike, MSEE,**
Principal Research Scientist, iCAD Inc.

**Jeffrey Hoffmeister, M.D. MSEE,**
VP & Medical Director, iCAD Inc.

### Key Takeaways

- ProFound AI Risk in breast cancer screening addresses women who may develop interval cancer and later stage cancer despite regularly attending breast cancer screening
- ProFound AI Risk is performed in women who do not have breast cancer detected from a screening mammogram and estimates the risk of breast cancer before or at next screening
- ProFound AI Risk identifies women who may benefit from precision screening, and therefore may experience earlier cancer detection, a lower treatment cost and potentially improved survival
- ProFound AI Risk has been developed for both Full Field Digital Mammography (FFDM) and Digital Breast Tomosynthesis (DBT) and offers an unprecedented level of short-term risk insight using several factors unique to the individual
- Area Under the Curve (AUC) for ProFound AI Risk DBT is 0.80, 18 points higher than traditional models even when factoring in breast density[9]
- ProFound AI Risk is designed for radiologists in the mammography screening workflow and is easy to use
- ProFound AI Risk is calibrated to 17 country specific breast cancer incidence and mortality rates
- ProFound AI Risk personalized risk information assists the radiologist in recommending tailored future screenings

## Introduction

### Motivation for Improving Risk Assessment

Traditional age-based screening mammography has been shown to reduce breast cancer mortality by 25-40% [1]. However, missed cancers, so called interval cancers, do occur [2]. Breast cancer is one of the most common cancers in women and is also one of the most common causes of death in middle aged women [3]. Women with a family history of breast cancer and/or a mutation in any of the established susceptibility genes, have an increased risk, but the majority of breast cancers are diagnosed in women without any history of the established risk factors [4-6]. There exists a need to estimate a woman's risk in developing breast cancer to mitigate missed and later stage diagnosis.

A risk-based screening approach to the detection of breast cancer is a challenging endeavor but understood to be the next evolution for precision medicine in breast cancer care. Models for estimating a woman's risk in developing breast cancer have been in use for decades [7], and tremendous work has been undertaken to use a woman's family history, lifestyle, and/or breast density to assess her individual risk [8]. However, the success of these models has been largely seen in subgroups rather than individuals. Artificial Intelligence (AI) based models are well poised to take advantage of the abundant information within the mammogram itself to enable far more accurate, personalized risk estimates to be generated.

Incorporating a risk-based metric, based on routine mammographic screening analyzed by an AI algorithm in tandem with multiple personalized factors, is an effective way to predict likelihood of cancer development in the near term. Such a risk model is expected to more accurately determine a woman's

short-term risk of developing breast cancer compared to traditional models. Additionally, it is expected to increase effectiveness of the screening process, increase efficiency, and simplify workflow for the clinician, improve patient experience, and reduce overall cost of care. Depending on the patient risk outcome, routine screening frequencies could be tailored (either reduced or increased) or supplemental screening may be recommended to better serve the individual patient needs.

iCAD Inc., in partnership with researchers from the Karolinska Institute in Sweden, has developed ProFound AI Risk for use with Full Field Digital Mammography (FFDM) and Digital Breast Tomosynthesis (DBT) mammography systems. ProFound AI Risk is the world's first individualized image-based model that incorporates racial, geographic characteristics, and a woman's age to estimate her individual risk for developing breast cancer. ProFound AI Risk is designed as a clinical support tool used to assess individualistic mammographic features such as density, texture, asymmetries, and shapes and patterns along with age. Using available incidence and mortality data from numerous countries, the incorporation of age, race, and geographic location permit ProFound AI Risk to provide a more individualized risk since incidence rates can differ significantly based on age and location. Using this combined information, Profound AI Risk generates a 2-year risk score for FFDM or a 1-year risk score for DBT which further stratifies women not already diagnosed with breast cancer into 4 risk categories (i.e., High, Moderate, General, and Low). This value provides the likelihood of a woman to develop breast cancer within 2 years for FFDM and within 1 year for DBT from her last normal mammogram, and her absolute risk (with corresponding risk category) can be used to guide future screening options/protocols. While the 2-year risk is the default output of ProFound AI Risk FFDM and 1-year risk is the default output of ProFound AI Risk DBT, iCAD has available extrapolation for varying intervals. As many countries have differing intervals of recommended mammographic screening, a 1-, 2-, or 3-year risk score is possible to complement US, EU, or UK screening protocols.

## Clinical Evaluation of ProFound AI Risk

### ProFound AI Risk FFDM

iCAD Inc., a pioneer in Artificial Intelligence (AI) algorithms using deep convolutional neural networks for breast cancer detection, has partnered with researchers from Sweden's Karolinska Institute who have developed a breast cancer risk model to assess short-term risk in parallel with AI analysis of 2D digital mammograms [9]. This model formed the basis for what is now ProFound Risk FFDM. The initial model (developed from a cohort of 974 cancers and 9376 healthy women) assessed risk based on analysis of the mammogram itself and incorporated factors of density, masses and microcalcifications, the asymmetry between left and right breasts, and age to generate a risk score. The advantage of ProFound AI Risk FFDM when compared to other models is its simplicity; it relies only on an assessment of the mammogram to derive its information and generate the risk score. No further information on family history of breast cancer and individual lifestyle such as Body Mass Index (BMI), menopausal status, alcohol/tobacco use, hormone replacement usage or genetic predisposition knowledge is required. ProFound AI Risk FFDM research shows that leveraging these features improves the overall performance for short-term risk assessment. However, the increase is subtle, and the superior performance of this model based solely on mammographic information is regarded as the best option.[9]

Assessment of models such as Tyrer-Cuzick and Gail are performed using the Area Under the Curve (AUC) from the Receiver Operating Characteristic (ROC) curve. Here, the sensitivity versus the false positive rate (defined as 1-specificity) is plotted and the area under the generated curve is calculated, where a value closer to 1 is understood to be ideal. Relatively speaking, an AUC of 0.5 (falling on the 45-degree line of the ROC curve) demonstrates no discrimination. By example, such an AUC can be interpreted as a hypothetical model consisting of two women, one of whom will develop breast cancer. With a 0.5 AUC, this model would identify the correct woman half the time; flipping a coin is equally effective. While there are varying levels of acceptability as the AUC approaches unity [10-11], doubt exists whether AUC's of approximately 0.60 are clinically useful.

The AUC results of the ProFound AI Risk FFDM model for a 2-year cancer risk are presented in Table 1 alongside the traditional Tyrer-Cuzick and Gail models (all having been run on the same dataset). The ProFound AI Risk FFDM model demonstrates an 11- and 12-point increase in AUC compared to the Tyrer-Cuzick and Gail models even when they are both augmented by breast density factors. To further establish the validity of the AUC determined by the ProFound AI Risk FFDM, three

external and independent datasets (consisting of 104, 613 and 179 incident breast cancer cases and random samples of 9,745, 8,489 and 9,491 healthy women, respectively) were utilized. For all three datasets, the determined AUC closely matched the originally determined value [9]. This research demonstrates the viability of ProFound AI Risk near-term risk information as a complement to long-term risk information in tailored screening.

*Table 1: AUC comparison of ProFound AI Risk FFDM model to traditional risk models with 95% confidence intervals provided [9]*

| Model | AUC (95% CI) |
|---|---|
| ProFound AI Risk FFDM | 0.73 (0.71, 0.74) |
| Tyrer-Cuzick* | 0.58 (0.56, 0.60) |
| Tyrer-Cuzick with density* | 0.62 (0.60, 0.64) |
| Gail* | 0.56 (0.54 ,0.58) |
| Gail with density* | 0.61 (0.60, 0.63) |

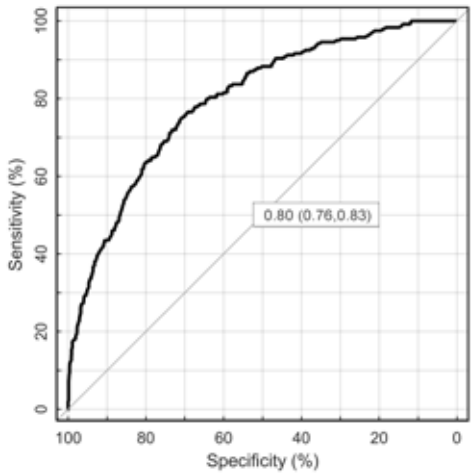
*Models were adjusted to yield a risk score for a 2-year interval

## ProFound AI Risk DBT

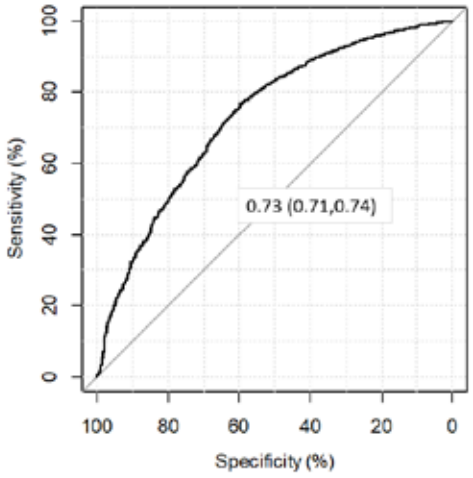
Building upon the success of ProFound AI Risk FFDM and its large datasets, an additional dataset was used to generate the ProFound AI Risk DBT product. A data set of 563 cancers ("cases") and 3,609 non-cancers ("controls") from four sites using three DBT manufacturers was used to train the algorithm. In this data set, the average age of all cases and controls was 59 ± 10 years with a one-year interval between DBT screening exams. An additional test set having an average age of 58 ± 10 years consisting of 240 cancers and 1,551 non-cancers obtained on three manufacturer's DBT systems from four locations that screened women annually was used to test the algorithm. The cases were prior screening DBT exams of women who later developed cancer at the next screening or as an interval cancer before the next screening, while the controls were screening DBT exams of women who were confirmed to remain healthy at the next screening. The DBT data were analyzed retrospectively with ProFound AI Risk DBT in order to assess the ability of the algorithm to estimate the risk of developing breast cancer in the interval before or at subsequent screening exams versus those remaining healthy.

In the study, ProFound AI Risk DBT returned a 1-year AUC of 0.80 which exceeds the 2-year AUC of 0.73 achieved on the ProFound AI Risk FFDM algorithm by 7 points and the modified Tyrer-Cuzick model by 18 points (Figure 1).

**ProFound AI Risk DBT** [19]




**ProFound AI Risk FFDM** [9]




**Tyrer-Cuzick plus density** [9]

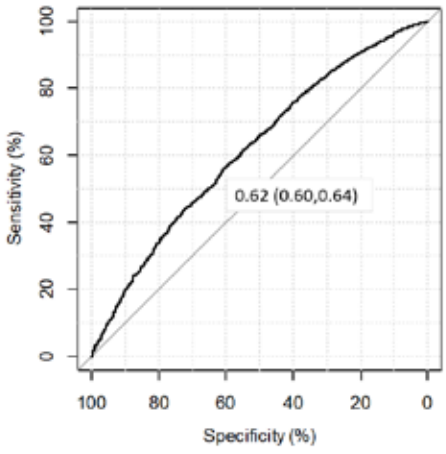



*Figure 1: Comparative ROC curves for ProFound AI Risk DBT and FFDM along with the modified Tyrer-Cuzick risk model. The AUC along with the 95% confidence interval boundaries for each model are presented.[9,19]*

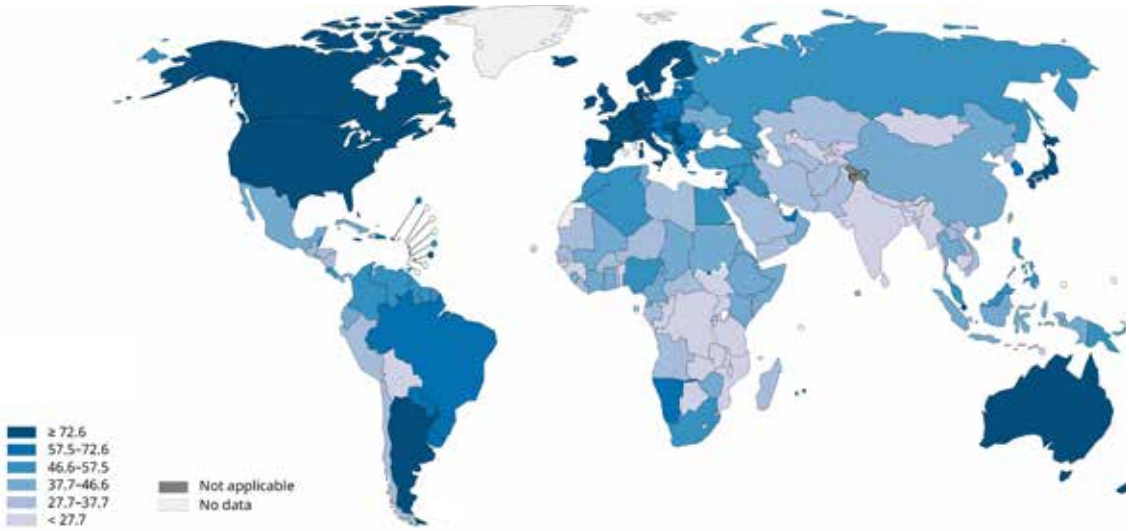



*Figure 2: Age standardized incidence rates for breast cancer worldwide. Values are in cases per 100,000. Data retrieved from: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf*

Integral to ProFound AI Risk DBT and FFDM is the inclusion of country specific estimates. The age-based incidence levels and mortality data from 17 countries (Table 2) enable more personalized risk calculations for women. The countries included coincide with the highest incidence of breast cancer as seen in Figure 2. North America, Western and Northern Europe, and portions of East Asia and Oceania constitute the highest incidences of breast cancer in the world [12]. By incorporating specific data from these regions into the ProFound AI Risk product, more effective prediction tools can be developed.

| Incidence and Mortality Countries | | |
|---|---|---|
| Australia | Germany | Spain |
| Belgium | Israel | Sweden |
| Canada | Italy | Switzerland |
| Denmark | Japan | United Kingdom |
| Finland | Netherlands | United States |
| France | Norway | Rest of World |

*Table 2: Countries with incidence and mortality data included with ProFound AI Risk*

The data from these 17 countries categorize a woman's likelihood of developing breast cancer by region, age, and within the United States, by different racial backgrounds as well [13].

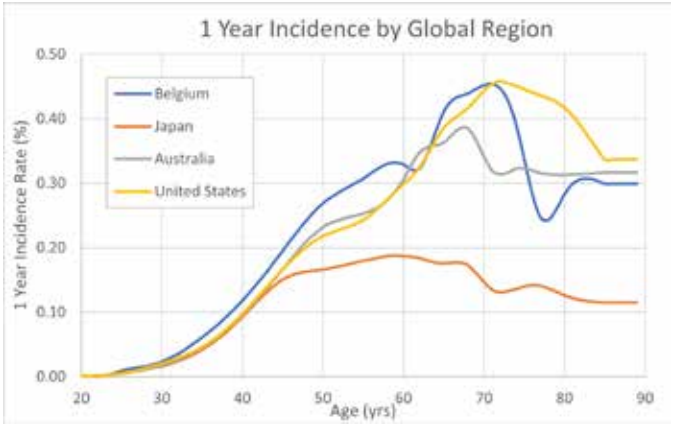



*Figure 3: Age-based breast cancer incidence and mortality data for world geographic regions*

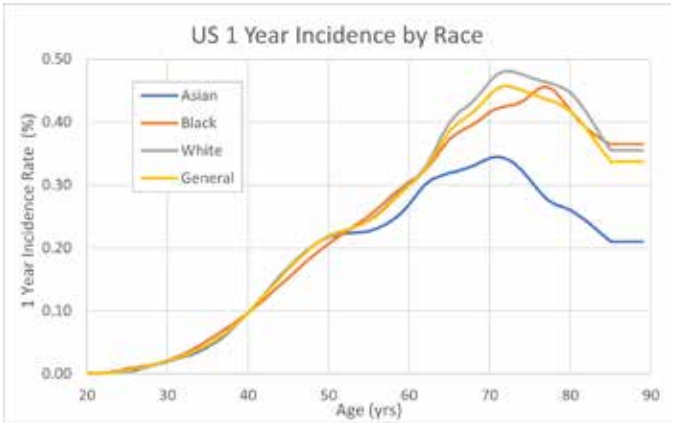



*Figure 4: Age-based breast cancer incidence and mortality data by race for the United States*

Figures 3 and 4 highlight the criticality of accurately predicting risk given variables of age, location, and race. It is clear from the plots that all are critical factors in

breast cancer risk and must be taken into account in any risk model. Depending on location, the maximum 1-year incidence rate will occur at different ages and the general trend and magnitude of the incidence as a function of age can vary significantly. Compounding the issue is race combined with geography; an Asian woman in the United States does not have the same age-based risk as an Asian woman in Japan. Equipped with such specific input details regarding geographic, age, and racial data, the ProFound AI Risk algorithm becomes better tailored to an individual.

## Clinical Implementation Suggestions

Several entities both in the United States, such as the United States Preventive Service Task Force (USPSTF) and the National Comprehensive Cancer Network (NCCN), and the United Kingdom's National Institute for Health and Care Excellence (NICE) have categorized breast cancer risk for clinical use [14-15]. Both stratify risk into the three groups (General, Moderate, High) which iCAD has expanded to four groups through the addition of a Low-risk category (Table 3).

Using Profound AI Risk DBT and the USPSTF/NCCN categorization, 44% of all screened women are considered to have a low risk of breast cancer (Table 3). In this group, not even one woman out of a 1000 will be diagnosed with breast cancer in the year or at next screening.

In contrast, the 15% of all screened women that fall into the high risk category have 16 times the risk of breast cancer compared the low risk population. Out of 1000 screened women in this high risk group, 9.8 women will be diagnosed with breast cancer within the next year or at the next screen. Correspondingly, for the NICE thresholds, 29% of the women align with the lowest risk level while 10% are in the highest risk category. Here, women at high risk have 28 times the likelihood to develop breast cancer within the next year or at the next screen than those at the lowest risk level.

To effectively measure the performance of ProFound AI Risk and compare its results to other models, an upcoming simulation is under investigation [16] in which the same screening cohort used for the development of ProFound AI Risk DBT was extrapolated to two other models (ProFound AI Risk FFDM and Tyrer-Cuzick with density). Analysis of the simulation data with respect to the risk thresholds provides an understanding of how effective one model may be in detecting cancers. More specifically, within the bounds of the simulation, if a certain proportion of higher risk women are selected for supplemental screening (assuming 100% sensitivity of the examination), the proportion of cancers having the potential to be detected can be determined.

New guidance from the American College of Radiology and the American Cancer Society points toward supplemental screening recommendations for women having a lifetime risk ≥20% [17-18]. According to the simulation cohort data, this threshold accounts for 12% of the women. Assuming this percentage of women is recommended for supplemental screening, the simulation indicates that the Tyrer-Cuzick with density model can identify up to 24% of the cancers (within a 1-year timeframe). By comparison, the value for ProFound AI Risk FFDM is 39% and for ProFound AI Risk DBT 47%. [19] Through this comparison, the superiority of the ProFound Risk products is clearly evident with nearly twice the ability relative to Tyrer-Cuzick model when addressing the near-term.

| | Stratification (Absolute Risk) | Women at Risk (%) | Average Absolute Risk (%) | Relative Risk |
|---|---|---|---|---|
| USPSTF /NCCN | Low (<0.12%) | 44 | 0.06 | 1.0 (ref) |
| | General (0.12% - <0.34%) | 28 | 0.21 | 3.5x |
| | Moderate (0.34% - <0.6%) | 13 | 0.45 | 7.5x |
| | High (≥0.6%) | 15 | 0.98 | 16.3x |
| NICE | Low (<0.075%) | 29 | 0.04 | 1.0 (ref) |
| | General (0.075% - <0.3%) | 40 | 0.16 | 4.0x |
| | Moderate (0.3% - <0.8%) | 21 | 0.49 | 12.3x |
| | High (≥0.8%) | 10 | 1.13 | 28.3x |

*Table 3 – Stratification of women using the ProFound AI Risk DBT and the USPSTF/NCCN-rooted and NICE guidelines. Risk categories, women at risk, and average absolute risk, and relative risk are described. [9]*

ProFound AI Risk is a clinical decision support tool that utilizes ProFound AI breast feature complexity findings, automated breast density, and age to calculate a woman's short-term, absolute risk of breast cancer. The output of ProFound AI Risk is a score in the form of a percentage of developing cancer in a specific timeframe. In the screen capture below (Figure 5) the 1-year risk, measured by ProFound AI Risk DBT, is estimated and described as "General" in a woman aged 45 years. The average risk of a 45-year-old woman is presented as 0.21% and the risk of breast cancer for this particular woman is 0.24%. This means that this woman has a 2.4 per 1000 chance of developing breast cancer within the coming year.


*Figure 5: Screen output of ProFound AI Risk DBT highlighting overall risk category, risk score, and alignment with population risk for the age of the woman screened*

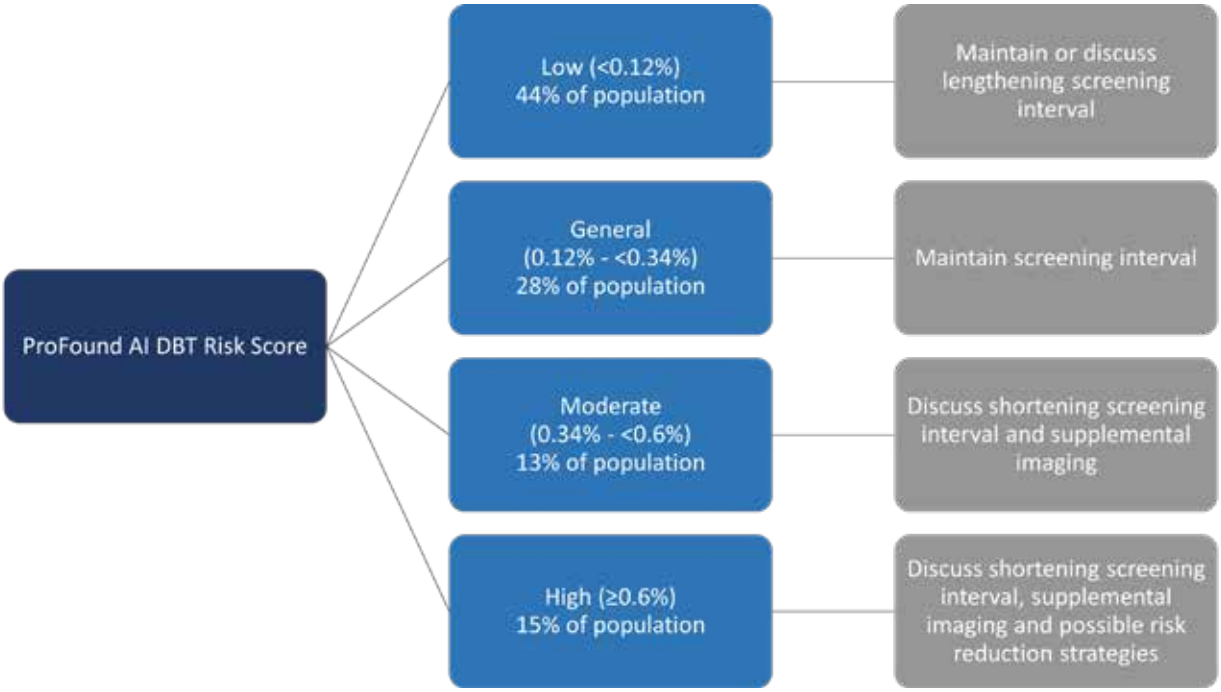
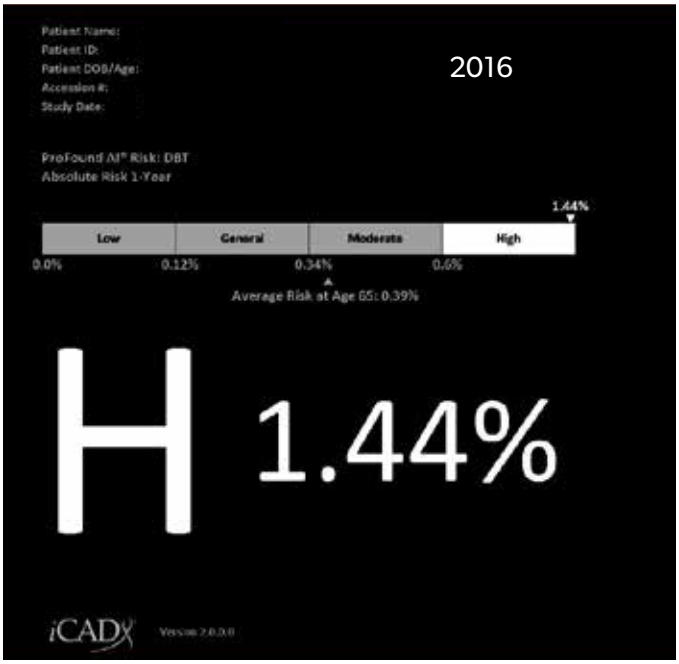
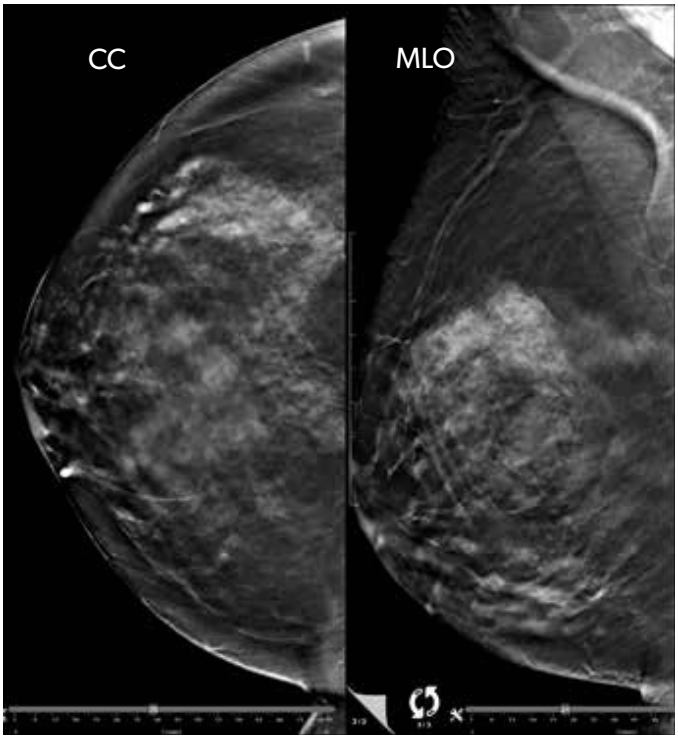



*Figure 6: The imagined future of screening risk stratifications for ProFound AI Risk DBT and possible screening recommendations depending on risk score, based on U.S. derived thresholds and a 1-year interval.*

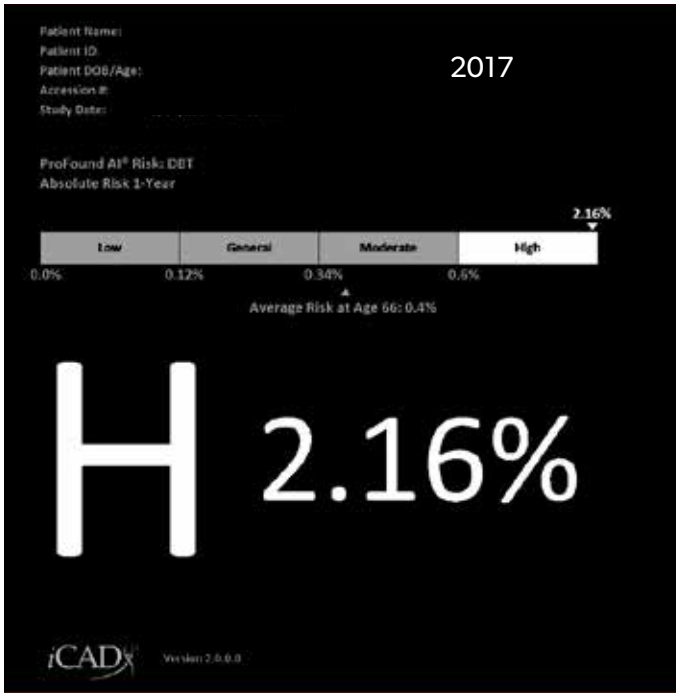
*a)*




*b)*




*c)*




*d)*

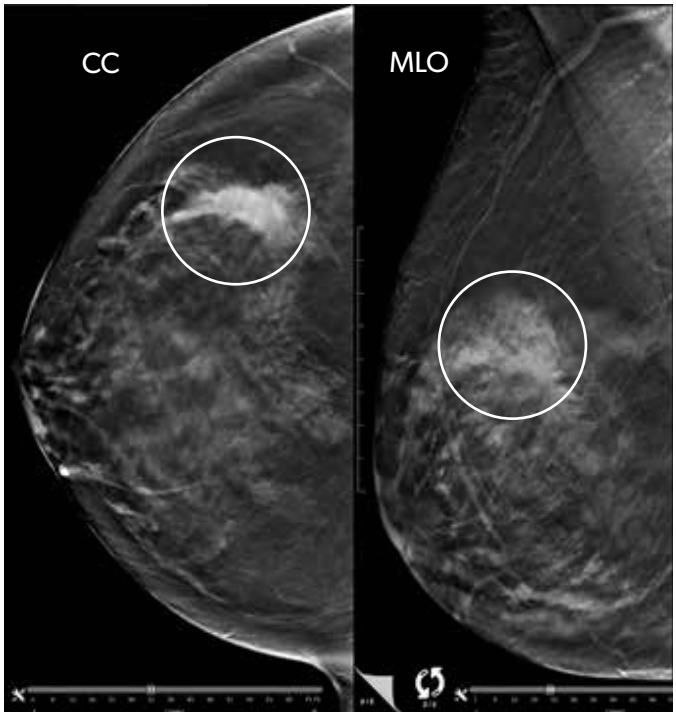



*Figure 7: Clinical example of ProFound AI Risk DBT, based on U.S. derived thresholds and a 1-year interval. Initial risk score result (a), and corresponding images (b) along with the following year risk score (c) and mammographic images (d)*

Based on clinical evidence and the risk score generated, future screening options can be customized for each patient (Figure 6). If a low risk score is presented, continuance of the woman's current surveillance plan or lengthened intervals between screenings at the physician's discretion, aligned with regional or national standards may be recommended. For General and Moderate risk categories, continuance of the current screening plan is likely to be recommended, but supplemental screening may be discussed. However, for women with high risk scores, recommendations may include shortening current screening frequencies. Additionally, supplemental screening options such as ultrasound, magnetic resonance imaging or contrast enhanced mammography coupled with genetic counseling, genetic testing, or risk reduction strategies may be discussed particularly with women with exceedingly high risk scores.

Consider the following example of ProFound AI Risk in a clinical scenario. A 65-year-old woman with a family history of breast cancer (maternal aunt was diagnosed with breast cancer at age 60) presented for her yearly screening mammogram in 2016. The DBT results indicated there was no evidence of carcinoma and the overall results were consistent with her prior mammograms. When ProFound AI Risk DBT was performed retrospectively on this mammogram, it returned a 1-year risk score of 1.44%, far above the threshold between the Moderate and High categories (0.6%), and nearly a factor of four times higher than the 0.39% average risk for a woman her age (Figure 7a). If ProFound AI Risk had been in use clinically, a risk score of this magnitude could have resulted in various screening options being presented immediately for earlier detection.

The following year, her normally scheduled DBT exam revealed an asymmetry in the lateral aspect of the right breast on the craniocaudal (CC) view, possibly corresponding to an architectural distortion in the mediolateral oblique (MLO) view at 10 o'clock (Figure 8d, as circled in the CC and MLO views), which was also palpable. A diagnostic ultrasound was then conducted that showed an irregular hypoechoic mass corresponding to the mammographic findings. Ultrasound-guided core biopsy revealed infiltrating ductal carcinoma (nuclear grade 2). The 2017 DBT exam, with an even higher 1-year risk score of 2.16% (Figure 7c), shows a suspicious finding that is more conspicuous than in the 2016 DBT images (compare Figure 7b to 7d).

Had ProFound AI Risk been used in 2016, a potential to detect the cancer earlier at a more treatable stage could have occurred either through recommendations for increased screening frequency and/or supplemental screening options.

## Developments in ProFound AI Risk assessment

Currently, ProFound AI Risk for DBT incorporates the same input factors as ProFound AI Risk FFDM; namely age and density factors originating from the mammogram itself. The decision to proceed with the incorporation of family history, lifestyle, and genetic markers is in product ideation phase by iCAD.

## Conclusion

ProFound AI Risk is designed to aid physicians in optimizing individualized screening in order to improve efficiencies and outcomes as well as reduce harms and costs. Whether FFDM or DBT, ProFound AI Risk offers unprecedented ability to estimate a woman's risk for developing breast cancer in a 3-, 2-, or 1-year interval. The sensitivity and selectivity of both exceeds traditional models by a large margin and offer a degree of personalization owing to the algorithm inputs for each woman. As the world's first short-term individualized image-based DBT and FFDM model incorporating racial and geographic features, ProFound AI Risk offers a great degree of customization. There is much potential to optimize patient screening with a personalized risk-based approach. A recent multi-site study further demonstrates the accuracy of ProFound AI Risk DBT, with similar performance to the iCAD internal data.[20] However, further clinical studies are required and are in process to optimize performance. Additionally, clinical efforts on a global scale are in motion to examine further applications and benefits for risk-based screening all in an effort to simply workflow for the clinician and overall patient experience.

## References


1. Lauby-Secretan B, Scoccianti C, Loomis D et al.; Breast-cancer screening--viewpoint of the IARC Working Group; N Engl J Med. 2015;372(24):2353-8.
2. Hoff S, Abrahamsen A, Samset J, et al. Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography—results from a retrospective review. Radiology 2012;264(2):378-386.
3. Malvezzi M, Carioli G, Bertuccio P et al. European cancer mortality predictions for the year 2019 with focus on breast cancer; Ann Oncol. 2019;30(5):781-787.
4. Neal CH, Rahman WT, Joe AI, et al. Harms of restrictive risk-based mammographic breast cancer screening. Am J Roentgenol. 2018;210(1):228-234.
5. Price ER, Keedy AW, Gidwaney R, et al. The potential impact of risk-based screening mammography in women 40–49 years old. Am J Roentgenol. 2015;205:1360–1364.
6. https://www.breastcancer.org/symptoms/understand_bc/statistics. Accessed July 2021
7. 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):1879-86.
8. Brentnall A, Harkness E, Astley S et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. Breast Cancer Res. 2015; 17:147.
9. Eriksson M, Czene K, Strand F et al. Identification of Women at High Risk of Breast Cancer Who Need Supplemental Screening. Radiology 2020; 297:327-333.
10. Mandrekar, J.N., Receiver Operating Characteristic Curve in Diagnostic Test Assessment, Journal of Thoracic Oncology, 2010; 5:1315.
11. Hosmer DW, Lemeshow S., Applied Logistic Regression, 2nd Ed. Chapter 5. New York, NY: John Wiley and Sons, 2000:160–164.
12. https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf Accessed July 2021
13. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2020 Sub (1975-2018) - Linked To County Attributes - Time Dependent (1990-2018) Income/Rurality, 1969-2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.
14. Melnikow J, Fenton JJ, Whitlock EP et al.; Supplemental Screening for Breast Cancer in Women with Dense Breasts: A Systematic Review for the U.S. Preventive Service Task Force [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2016 Report No.: 14-05201-EF-3.
15. National Collaborating Centre for Cancer (UK). Familial Breast Cancer: Classification and Care of People at Risk of Familial Breast Cancer and Management of Breast Cancer and Related Risks in People with a Family History of Breast Cancer. Cardiff (UK): National Collaborating Centre for Cancer (UK); 2013 Jun.
16. Eriksson M, Destounis S, Schilling K, A Risk Model for Digital Breast Tomosynthesis to Predict Breast Cancer and Guide Clinical Follow-up, (in review)
17. Lee C, Dershaw D, Kopans D et al. Breast Cancer Screening with Imaging: Recommendations from the Society of Breast Imaging and the ACR on the Use of Mammography, Breast MRI, Breast Ultrasound, and Other Technologies for the Detection of Clinically Occult Breast Cancer, J Am Coll Radiol 2010;7(1):18-27
18. Oeffinger K, Fontham E, Etzoni R, Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update from the American Cancer Society, JAMA 2015;15:1599-1614
19. iCAD data on file
20. Eriksson M, Destounis S, Czene K, et al. A risk model for digital breast tomosynthesis to predict breast cancer and guide clinical care. Sci Transl Med. Epub 11 May 2022. 14 (644). DOI: 10.1126/scitranslmed.abn3971