Quantitative MR Imaging Markers For Breast Cancer Risk Prediction

Wesley Surento

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2026

Reading Committee:
Savannah C. Partridge, Co-Chair
John H. Gennari, Co-Chair
Brian N. Dontchos
Christopher I. Li
Eardi Lila

Program Authorized to Offer Degree:
Department of Biomedical Informatics and Medical Education
University of Washington School of Medicine

University of Washington

Abstract

Quantitative MR Imaging Markers For Breast Cancer Risk Prediction

Wesley Surento

Co-Chairs of the Supervisory Committee

Savannah C. Partridge

Departments of Radiology and Bioengineering

John H. Gennari

Department of Biomedical Informatics and Medical Education

Breast cancer risk assessment can help early detection and intervention planning, lower mortality, and improve care outcomes. Many current tools utilize risk factors such as age, menopausal status, personal biopsy history, and family history of breast cancer. Although these characteristics are associated with risk, the tools incorporate minimal information from imaging studies such as mammograms or MRIs. Even when negative for disease, analysis of these images could provide rich information on breast tissue biology that might also capture risk. There remains much unexplored potential in leveraging imaging markers as part of risk assessment models for prevention and early prediction.

In this work, we aim to develop and assess a new risk prediction model that incorporates MRI-derived features along with clinical risk factors.

In Part 1, I collected relevant clinical risk factors for predicting cancer within 5 years from the MRI for a cohort of elevated risk patients. I built a REDcap data repository to store their clinical risk factors obtained from electronic health records, such as select demographic information, menopausal status, BRCA mutation status, and family history of breast cancer.

In Part 2, I identified predictive quantitative breast MRI markers. I extracted quantitative markers of background parenchymal enhancement using an image processing pipeline, which I helped develop, and assessed their ability to predict 5-year cancer risk.

In Part 3, I performed multimodal combination of imaging markers and clinical risk factors to predict 5-year cancer risk. I also explored the performance of these combinations for several cohort subsets. My findings showed that multimodal combination with imaging markers can improve the predictive performance of a conventional risk assessment tool.

Table of Contents

Dedication

To my family, current and future.
You are the reason I strive to become a better version of myself everyday.

Epigraph

“Essentially, all models are wrong, but some are useful”

-George Box (1919-2013) and Norman Draper (1931-2022)

Acknowledgments

A few lines here will not do justice to the full extent of how I want to acknowledge every single person mentioned. Whatever I managed to write here is merely what I could put together under the time constraints I had. These are not meant to fully characterize my relations with every single individual; they are but a brief snapshot in time.

I want to start by thanking my advisors, Savannah C. Partridge and John H. Gennari. They have always been generous to me with their time, resources, and patience. They chose to take a chance on me before knowing as well as they do now, and that gave me a real shot at doing any of this. I was able to much from them about becoming a rigorous scientist, and about becoming a more thoughtful and compassionate person. I am grateful to have been in their care during this journey that was approximately equal parts challenging and rewarding. Even as I continually grow and develop into a more complete individual, I intend to take the qualities I nurtured here to wherever I go in the future, hoping to be a positive influence to those that I will meet.

I thank the members of my dissertation committee, who have provided insightful directions to my dissertation work. I first met them through a variety of reasons; taking Eardi's class, knowing Brian through the Breast Imaging section in Radiology, and meeting Chris as Savannah's collaborator. I still recall the time in my third year when I reached out to each of them, pitching my dissertation aims using a brief presentation of slides; all of them said yes to me on the spot, much to my surprised relief. From that first meeting, to the General Exam and final defense, they offered questions which helped to shape and improve my work. Although I have not had the chance to work closer with them, I look forward to opportunities that allow me to collaborate with them again in the future.

I would like to acknowledge several close friends from my cohort, who have been faithful comrades-in-arms to me on this journey. We cannot take each other's place in our respective battles, but we knew that burden can always be shared and made lighter.

With Raghav I have shared much camaraderie and solidarity in matters of grad school and beyond; I just can't get enough of spending time with you. Now that we're both done with grad school, we likely have more time to explore common interests and hobbies, along with the occasional griping about life. You taught me how to swing a golf club for the first time and I'm just glad that nobody in the vicinity got hurt when that happened. I want to do so much more together; maybe we'll try pickleball and archery and anything we haven't thought of yet. I'd also like to taste more of each other's cooking. And of course, there is so much sports out there for us to watch together. Can't wait to do that and talk about them with you!

With Su I have spent the good part of my first four years of grad school; all the grocery store runs, anime watching, and finding new places to eat. I got to experience my first live concert and e-sports finals with you and I would love to go to more together. One day I may yet succeed in watching a game of baseball with you. I am glad that after the hard work of every quarter we always knew that satisfaction can be found in some phenomenal food at Maneki. Good times! Thank you for courageously agreeing to be the first taste-tester whenever I tried a new dish; I hope the long-term returns outweighed the initial risks for you. There is certainly much more to do in the future; we just have to make things happen.

With Velvin, I spent numerous fun weekends along with our group. I always looked forward to the weekend after a long week of work; spending time enjoying music and watching videos on

YouTube and Bilibili. Year 1 to 4 would be a lot harder without you and the group! I also want to mention that I am inspired by your work ethic and resilience; I am proud of what you've accomplished and overcome!

I want to thank my coworkers in the Quantitative Breast Imaging Lab. You are my work family and experience the day-to-day challenges and triumphs with me.
To Smita: Thank you for being a sounding board to my many thoughts and ideas. You are probably the first person in the lab that I became familiar with. Going to conferences with you is so fun; I wish I could do more!
To Isabella: I am so happy we became new baseball fans around the same time! I always have so much fun going to games with you and Yin. Hopefully the team does better every year!
To Anum: Thank you for your guidance and being patient with everything that I do.
To Habib: Thank you for being a source of support and encouragement! Thank you for the fun boat rides and showing us around Lake Union!
To Mike: Thank you for being a fun and patient cubicle mate!
To everyone else in QBI: Thank you for making QBI lab what it is! My journey here would not be the same without you!
To Yin: I am so glad I found a fellow baseball enthusiast in you! Being able to talk ball with you helps deepen my love for this game! I look forward to many more years of fun rooting for this glorious (and miserable) baseball team of ours! Go M's!

There are many others in BIME whom I have had the pleasure to interact with.
To Xiaoyi: Thank you for being such a reliable project teammate in Year One; I truly felt that working with you and Raghav showed me what a team looks like when we are more than the sum of our parts.
To Xinyang: I am really grateful for all the fun we had with the group on weekends; and your cooking never fails to surprise! Hope to visit you some time on the East Coast!
To Mu: Thank you for your continued encouragement and support to the meal gatherings! love all the stuff you bring back everytime you visit Taiwan. I also appreciate the conversations we had talking about everything from grad school to world news. I guess you were still unintentionally fulfilling Peer Mentor duties after all this time.
To Amber: Thank you for being kind and caring in every conversation and discussion. Your enthusiasm for knowledge continues to inspire me.
To Peter Ju: Man, I regret not starting to watch sports earlier; I bet we would've had a lot more fun together! Hope to visit you and watch a game together! Thank you for serving as Social Chair with me that one year!
To Faisal: Thank you for leading 591 with me in Autumn 2024! As I mentioned to you before, I am really impressed by how you can balance grad school and parenting all this time

To everyone who took 591 Intro to Selected Topics in Medical Imaging and Informatics in Autumn with me and Faisal (Joyanne Huang, Yein Jeon, Peter Ju, Kevin Li, Sitong Liu, Zoljargal Lkhagvajav, Raghav Madan, Ojas Ramwala, Zhaoyi Sun, Yongsen Tan, Su Xian, Kenny Yi, Avery Yu, Sihang Zeng, Wenyu Zeng, Yujie Zhao, Sitong Zhou): A

big thank you for giving me the opportunity to learn from all of you. It was a busy but especially rewarding quarter for me.I am excited to see the bigger and better things that all of you will do. Best wishes in all that you do!

To Jim Phuong, Andrew Teng, Reggie Casanova-Perez: Thank you for showing me around places in Seattle and all the places with great food! Looking forward to hanging out with you more!

Lastly, I want to thank my family for their unending support and encouragement. I would not be here without you, and I hope to continually become better for you. There aren’t enough words to express my gratitude, so I hope to make up for that through my time with you.

This dissertation work is supported by NIH/NCI grants U01CA152637, R01CA207290.

Chapter 1
Introduction

1.1 Motivation

Breast cancer accounts for an estimated 32% of newly diagnosed cancers annually for women in the United States, and is the second leading cause of cancer related deaths for them (Siegel et al., 2025). Approximately one in eight women in the US will develop breast cancer in her lifetime (Giaquinto et al., 2022), representing a significant health risk for every individual in the long-term. Screening and early detection of breast cancer lead to better care outcomes (Duggan et al., 2021; Hoerger et al., 2011), including for recurrent cancer cases (Houssami et al., 2009; Lu et al., 2009). A study investigating breast cancer mortality between 1975 and 2019 attributed 25% of mortality reduction to screening and early detection (Caswell-Jin et al., 2024).

Early detection of breast cancer is achieved through a combination of the use of breast cancer risk assessment models, which estimate a risk score based on personal and familial risk factors (Jacobi et al., 2009; McClintock et al., 2020; Waters et al., 2020), and radiological screening (Lee et al., 2020; Monticciolo et al., 2017). Risk assessment tools and imaging occupy different parts in that sequential workflow, but having good performance in both is crucial for ensuring overall performance of the entire screening process.

First, an individual can be classified as being at average or high risk of developing cancer in a variety of ways, which include but are not limited to: lifetime risk calculated from assessment models, personal risk factors or family history. Lifetime risk scores from risk assessment tools can be useful in more personalized risk stratified screening strategies. Most risk assessment tools utilize clinical risk factors such as age, prior breast biopsy history, and family history of breast cancer (Cintolo-Gonzalez et al., 2017a).

Recommendations for screening modality and frequency are based on a woman's estimated risk, where average risk women in the United States are typically recommended to undergo annual screening with mammography and high-risk women are recommended to receive additional screening using magnetic resonance imaging (MRI), two commonly used imaging modalities for breast cancer screening. The sensitivity of radiological screening technologies is also of high importance for early detection. In high-risk populations, MRI is reported to be more sensitive (sensitivity range: 75% to 91%) than mammography (sensitivity range: 36% to 40%) but may sometimes be less specific (specificity range: 82% to 93%) than mammography (specificity range: 93% to 99.5%) at detecting first-time cancer (Kriege et al., 2004; Warner et al., 2008).

Accuracy in both the risk assessment and radiological screening stages is crucial for the overall screening. There are consequences to inaccuracies in the entire screening process. False positives result in psychosocial (Brodersen & Siersma, 2013) and financial (Elmore et al., 1998) burden to women, while false negatives result in delayed diagnosis and intervention as well as financial costs (Petticrew et al., 2001). Therefore, there is need to improve current approaches of risk assessment in order to select optimal screening regimens for early detection in order to provide timely intervention and enable better care outcomes. This work aspires to improve risk assessment tools during the initial part of the screening process, which we will discuss further in the next section.

1.2 What this work aims to do

This dissertation aims to develop and assess a new risk prediction model that uses a multimodal combination of MRI-derived features and clinical risk factors. The central hypothesis of this dissertation is that combining imaging and clinical risk factors will produce better prediction performance than using either alone. Since only a few risk assessment tools currently incorporate mammographic breast density as an imaging factor (Cintolo-Gonzalez et al., 2017a; Ready et al., 2010), there is value in investigating the predictive performance of markers derived from other imaging modalities, particularly MRI. Conventional risk models use clinical factors that are common and predictive at the population level but lack individual-level information about the person's tissue biology; age is a common characteristic, but tissue biology is unique. Imaging can provide patient-specific information that better reflects the individual's own biology at a particular point in time. Therefore, there is an opportunity to improve cancer prediction tools that use only clinical risk factors through incorporating imaging markers. While not all risk assessment tools incorporate information from imaging, those that incorporate mammographic density report an improvement of area under the receiver operating curve (AUC) of 0.04, from 0.57 to 0.61 for the Tyrer-Cuzick model and 0.55 to 0.59 for the Gail model (Brentnall et al., 2015). The differences between these models will be discussed further in Chapter 2. While the increases in AUC after adding mammographic density are relatively modest, they do demonstrate the value that imaging markers potentially hold in improving risk models. Overall, the performance of such clinical risk models remain modest (0.59 to 0.61), suggesting room for improvement. This dissertation work leverages functional MR imaging markers, capturing tissue physiology, to further improve performance of an established clinical cancer risk prediction model for a population of women considered to be at high risk of developing cancer. The proposed multimodal risk prediction model was aimed at improving upon the population-level clinical risk factors by using the individual-level features of quantitative MRI markers.

The standard screening modality for women at average risk is mammography, while high-risk women receive MRI in addition to mammography during routine screening (Smith et al., 2019). As mentioned before, in the high-risk screening population, MRI is more sensitive than mammography (Warner et al., 2008). Another important aspect of

difference between the two imaging modalities is cost; a study estimated a bilateral mammogram to cost between $33 to $73 while MRI could be between $646 to $1432 (Moore et al., 2009), which is many times more expensive. This is another reason why MRI is offered only to individuals who will benefit the most from getting it; women who are considered to be at high risk of developing cancer.

The specific population we study in this work is a subset of high-risk women who underwent breast MRI screening. Individuals are considered to be at elevated risk of developing breast cancer due to a combination of family history or personal risk factors. Our work aims to better stratify women considered to be at high risk and enable better personalization of screening and intervention strategies in contrast to a one-size-fits-all approach for this risk group. If the risk prediction model from our work is shown to improve upon conventional risk assessment tools, there is opportunity for a future iteration of the model to provide complementary information to guide in the clinical decision-making process for adjusting screening strategies for high-risk individuals. In Chapter 6, I discuss the implications that this research would have on the development of a clinical decision support tool, even though it is beyond the scope of this dissertation. For individuals who are identified to be at the lower end of the high-risk spectrum, there can be opportunities to de-escalate screening strategies to biannual MRI instead of an annual frequency. For individuals who are identified to be at the higher end of the high-risk spectrum, there can be opportunities for escalation which may include but not limited to additional preventive therapies such as aromatase inhibitors or selective estrogen receptor modulators.

The next section provides an outline of my work, beginning with the three Parts which roughly guided the chronological order of work done, then followed by a summary of every chapter in this dissertation document.

1.3 Three Parts and Outline of this dissertation

This dissertation work was conducted in three sequential parts:

- Part 1: Identified a study cohort and extracted relevant risk factors for predicting future breast cancer
- Part 2: Identified predictive quantitative breast imaging markers from screening MRIs
- Part 3: Developed a new multimodal cancer risk prediction model

To create and evaluate a new multimodal risk model leveraging MR imaging markers, I first needed to identify and collect data from a specific study cohort, which is the work of Part 1. Then, in Part 2, I calculated MR imaging markers and assessed their performance in predicting cancer risk. In Part 3, I showed that multimodal combination of

imaging and clinical risk scores perform better than using either alone for predicting cancer diagnosis within 5 years.

Chapter 2: Imaging in cancer screening and breast cancer risk prediction models

Before I delve into the methodology and results of this work, it is necessary to provide key context. I begin with discussing the role of imaging in breast cancer screening, looking at various modalities and eventually focusing on MRI since it is most relevant to this work. I also briefly discuss screening recommendations in average- versus high-risk women. This is followed by a discussion of common imaging markers. Since this dissertation is centered around MRI in the high-risk population, much of the work will explore qualitative and quantitative MRI background parenchymal enhancement (BPE).

I examine various breast cancer risk assessment tools and their differences. I inspect their different risk factors and examples where they have performed well in risk prediction. I eventually highlight the Tyrer-Cuzick risk model and why we select it as a benchmark for comparison in our work. In the last part of this chapter, I discuss emerging imaging-based risk models. They are different from the approach taken in this work, while still relevant to mention as newer examples of cancer risk prediction.

Chapter 3: Preliminary work on quantitative MR markers associated with cancer risk

This chapter contains our preliminary work in exploring the association between quantitative MR markers and breast cancer risk. The patient cohort included 77 women with dense breasts and varying breast cancer risk who underwent breast MRI. This cohort did not yet have 5-year cancer outcome and follow-up available. In lieu of true outcomes of cancer diagnoses, Tyrer-Cuzick risk scores were available and binarized into high- and low-risk categories. This work was published as a full-length article in *Cancers* (Surento et al., 2025). Through exploring various quantitative imaging markers, we found that quantitative BPE metrics were most highly correlated with Tyrer-Cuzick risk groups, encouraging us to investigate their performance in a larger cohort with actual cancer outcomes.

Chapter 4: A data repository for the study cohort

This chapter describes the data we collected for the study cohort as Part 1 of this dissertation. I begin by defining our study cohort, detailing the process by which we apply exclusion criteria and case-cohort sampling strategy to decrease the total cohort size to a more feasible number to extract quantitative imaging marker and clinical risk factor lookup. The chapter presents the motivation and details of this sampling approach, and we arrive at a cohort of 234 women.

I then describe building a data repository for the study cohort, enumerating the various clinical risk factors and imaging markers to be collected, along with their data sources. Much of the information for clinical risk factors is contained in medical free text, for which I conducted manual lookups of medical records and read through physician notes to extract the relevant data. I stored and organized the resulting data in a REDcap data repository. Using the clinical risk factors, I then calculated a risk score using the Tyrer-Cuzick model, introduced in Chapter 2. The approached described here for risk factor collection and management may serve as a reference guide for future multimodal work that combines imaging and clinical factors.

Chapter 5: Quantitative MR imaging markers and cancer risk

This chapter assesses the predictive value of the quantitative imaging markers developed and evaluated in Part 2 of this dissertation. The primary objective is to determine whether MRI-derived quantitative imaging markers can predict a future cancer diagnosis. Our final study cohort of 234 women, as described in Chapter 4, all have 5-year cancer outcomes available. In this chapter, we assess the relative performance of predicting cancer diagnosis within 5 years using qualitative and quantitative BPE. Here, we found that quantitative BPE metrics significantly predict future breast cancer diagnosis and performed equal to or better than qualitative BPE in predicting 5-year cancer.

Chapter 6: A new imaging-based cancer risk prediction model

This chapter presents the approach and results of this dissertation's main goal: building and evaluating a multimodal risk prediction model for predicting 5-year cancer risk. As Part 3 of the dissertation work, this chapter examines the predictive performance of models that combine individual quantitative BPE metrics with Tyrer-Cuzick risk scores derived from the clinical risk factors described in Chapter 4.

We found that in the full cohort of 234 individuals, multimodal prediction generally performed better than either clinical or image-based prediction. We also explored model performance in several subgroups: 182 ndividuals without known BRCA mutations (i.e. excluding BRCA-positive women), 212 individuals who developed invasive cancers (excluding those with ductal carcinoma in situ [DCIS] only), and 138 individuals who were pre-menopausal. These subset analyses provided insights into relative model performance in these populations.

Chapter 7: Conclusion and Future Work

The concluding chapter begins with addressing limitations of our current work and discussing potential avenues for improvement. An important limitation was the missingness of clinical risk factors, where individuals had varying levels of data completeness.

In addition, I present a road map for the next steps of the project. Short-term next steps include expanding the sample size through expanding the retrospective study's follow-up period to a later end-date and trying alternative prediction models or performance metrics to identify optimal combinations of variables. Long-term next steps for the project may involve refining the variables used for prediction. For quantitative imaging markers, there may be alternative approaches to derive quantitative metrics such as selecting different enhancement thresholds and acquisition time points along the dynamic contrast-enhanced image series. For clinical risk factors, there may also be efforts directed at collecting additional risk factors not included in our database. Such alternative approaches may ultimately produce different optimal combinations of imaging and clinical risk variables, leading to new insights and best practices.

This chapter also discusses broader implications and directions for future work in multimodal risk prediction for cancer and concludes with a brief summary of the clinical and biomedical informatics contributions from this dissertation work. Overall, the results support the multimodal approach of supplementing conventional clinical risk prediction with imaging markers.

1.4 Contributions of this dissertation work

The results from our work contribute to a growing body of evidence demonstrating the value of quantitative markers in predicting breast cancer risk. We generally observed automated quantitative BPE metrics to perform just as well or better than qualitative, radiologist assessed BPE. As quantitative metrics have the potential to be more objective and reproducible markers of risk than qualitative BPE, our results offer supporting evidence to incorporate quantitative BPE in future cancer risk prediction.

An important contribution from my work is validating that multimodal prediction approaches outperform those relying solely on clinical factors or imaging markers. Despite the challenges associated with collating imaging and clinical risk data from disparate sources, this work has demonstrated the inherent value in improving prediction performance through such combinations. Imaging features offer a more direct reflection of an individual's biology, while clinical risk factors offer a further view of disease likelihood as recorded and followed at a population level. Prediction approaches that combine data capturing both aspects may enable a more balanced and holistic risk assessment compared to using either separately. While future work could explore more optimal

combinations of imaging and risk factors, our work provides insights and results as an early proof-of-concept.

While beyond the scope of this dissertation, insights from a future iteration of our multimodal risk prediction may potentially assist in clinical decision support. Our work is partly motivated to help stratify patients who are categorized to be at high risk for developing cancer. When there is a future iteration of a multimodal risk model with improved performance, it may be used to categorize patients to be at the higher end versus lower end of the risk spectrum. If validated well, the predicted outcome may be used to recommend de-escalation or escalation of screening and interventions. De-escalation could be in the form of a reduction in frequency of MRI screening, or an escalation by prescribing risk-reducing medication such as aromatase inhibitors or selective estrogen receptor modulators.

In summary, my work contributes in several ways to research and potentially to clinical practice. I explored quantitative imaging markers for predicting cancer risk, I validated of multimodal risk prediction, and potentially provided an initial precursor to clinical decision support for planning an individual's screening and intervention.

Chapter 2

Imaging in cancer screening and breast cancer risk prediction models

2.1 Screening recommendations and introduction to imaging modalities

This chapter discusses two important aspects of breast cancer screening: imaging and risk assessment tools. In Sections 2.1 and 2,2, I discuss the imaging modalities and markers used during cancer screening, followed by Section 2.3 with a survey of various risk assessment tools used to calculate lifetime risk scores. Together, they form important elements of breast cancer screening.

As described in Chapter 1, breast cancer is a serious public health hazard affecting women in the United States, being the second leading cause of cancer related deaths and diagnosed in approximately one in eight women over their entire lifetime (Giaquinto et al., 2022). Screening and early detection of breast cancer can lead to better care outcomes (Duggan et al., 2021; Hoerger et al., 2011) and have been credited with a 25% mortality reduction between 1975 and 2019 (Caswell-Jin et al., 2024).

The American College of Radiology recommends annual mammographic screening to begin at age 40 for women at average risk for breast cancer, and age 30 for women determined to be at high risk due to factors including but not limited to: carrying a genetic mutation, family history, or a calculated lifetime risk score of greater than 20% (Lee et al., 2020; Monticciolo et al., 2017). An individual's lifetime risk score is calculated during screening by using risk assessment tools, which will be discussed in further detail in Section 2.3 of this chapter. High-risk women are also recommended to have MRI as an additional imaging modality for annual screening beginning between ages 25 to 30 years old (Lee et al., 2020). The high prevalence of elevated mammographic density among high-risk women renders mammography less effective at identifying cancer since dense breast tissue may potentially mask the presence of cancer already present at the time of screening, leading to underdiagnosis (Boyd et al., 2007; Porter et al., 2007).

Mammography and MRI are two commonly used imaging modalities for cancer screening, with mammography being more common because it is offered to both average-risk and high-risk groups of women while MRI is recommended as an additional modality just for the high-risk group due to its limited access and high cost. A mammogram is a

specialized type of X-ray imaging that captures the distribution of breast tissue composition: fatty tissue has a darker appearance while breast parenchyma including the ducts and lobules attenuates more X-ray beams and has a brighter appearance. A typical mammogram offers two views for each lateral breast: the craniocaudal and mediolateral oblique views.

Breast MRI screening involves three-dimensional scans of contrast-enhanced and non-contrast-enhanced image acquisitions. The BI-RADS Atlas recommends breast screening images to be acquired using pulse sequences that include T1- and T2-weighted images with or without fat saturation, dynamic contrast-enhanced (DCE) acquisitions involving multiple post-contrast time points (after injection of a gadolinium based contrast agent), along with subtraction and maximal intensity projection (MIP) post-processed images (Arian et al., 2022; D'Orsi et al., 2013). T1- and T2-weighted images provide anatomical information to help radiologists assess morphology, while DCE and MIP images show functional information involving intravenously injected contrast material enhances tissue surrounding vessels and helps radiologists assess physiology. On MRI non fat-suppressed T1-weighted images, breast parenchyma has high water and protein-bound cellular content and so appears relatively dark due to slower T1 relaxation rates, while fat appears bright due to more rapid T1 relaxation rate. On T2-weighted MR images with fat saturation, breast parenchyma appears darker because it contains less free water due to presence of cellular components and has shorter T2 relaxation time relative to pure fluid, while fat appears dark due to the fat saturation technique that suppresses signals from protons in fat molecules. DCE is performed using sequential fat-suppressed T1-weighted scans acquired before and multiple times after contrast injection (gadolinium based contrast agent). On DCE images, breast parenchyma shows varying degrees of enhancement depending on the amount of contrast agent uptake. MIP images depict the highest signal from across all breast slices in the three-dimensional volume, showing regions with the greatest contrast material concentration, including blood vessels and potentially suspicious breast lesions. This is a summary of what appears on breast mammography and MRI.

Apart from mammography and contrast-enhanced breast MRI, there are also other modalities for breast cancer screening such as ultrasound and contrast-enhanced mammography (CEM). For women with dense breast tissue, mammographic screening supplemented with ultrasound showed improved sensitivity in detecting cancer compared to just using mammography alone (Berg et al., n.d.; Giuliano & Giuliano, 2013). However, this is also accompanied by more false-positive findings when compared with mammography alone (Giuliano & Giuliano, 2013; Hooley et al., 2012). Contrast-enhanced mammography involves the injection of iodinated contrast agent two minutes prior to imaging, allowing better visualization of vascularity (Jochelson & Lobbes, 2021; Sogani et al., 2021). CEM has been found to detect lesions better than conventional mammography (Cheung et al., 2014; Fallenberg et al., 2017).

Despite the availability of various imaging modalities for breast cancer screening, MRI remains to be the best screening tool for individuals considered to be at high risk of developing cancer. Multiple studies comparing the sensitivity of MRI against CEM (Neeter et al., 2023; Pötsch et al., 2022) and ultrasound (Berg et al., 2012; Kuhl et al., 2010; Kuhl et al., 2005) have reported MRI outperforming other modalities. While false positive findings are present in MRI, studies involving multiple rounds of screening report decreases in false positives; exams interpreted as suspicious decreased from 26% in baseline screening to 10% in subsequent screening (Warner et al., n.d.), and callback rates decreased from 15% at first screening to 9% at subsequent rounds of screening. Thus, MRI is a modality that offers high sensitivity and shows reduced false-positive rates over multiple screening instances.

2.2 Imaging markers for screening

In mammographic screening, radiologists look for abnormal patterns in the breast tissue. Mammographic images enable detection of tumorous growth before they become palpable. Natural variations in every individual's tissue composition of fat, stroma and epithelium causes differences in radiographic appearance; fat appears dark on a mammogram while epithelium and stroma appear light (Johns & Yaffe, 1987). Deviations from healthy appearances are therefore used as indicators of abnormal growth; especially in detecting the subtle signs indicative of an early tumor. Radiologists are trained to look for these signs, such as clustered calcifications, poorly defined masses, dilated ducts, architectural distortion, asymmetry, and developing density (Sickles, 1984). A typical mammographic report will describe the presence or absence of these findings, along with a qualitative assessment of the amount of stroma and epithelium as one of four categories of mammographic density (MD): almost entirely fatty, scattered areas of fibroglandular density, heterogeneously dense, or extremely dense.

MD is always reported regardless of the presence or absence of lesions. This makes MD arguably the imaging marker with the largest availability in screening exams, by virtue of it being present in all average-risk women who are offered mammographic screening. An important study involving 1112 matched case-control pairs showed that higher MD is strongly associated with increased risk of breast cancer (Boyd et al., 2007). Naturally dense tissue can also potentially mask smaller lesions and prevent their detection during radiological interpretation. Aside from being a qualitative category, mammographic density may also expressed as a quantitative value using tools such as Cumulus (Byng et al., 1998), LIBRA (Keller et al., 2015), Volpara (Matakina Technology, Wellington, New Zealand) and Quantra (Hologic, Bedford, MA, USA).

When interpreting MRI screening scans, radiologists look for tissue characteristics and potential abnormalities on images from multiple pulse sequences. As mentioned earlier, an MRI breast screening is recommended by the BI-RADS Atlas to involve image acquisitions from several pulse sequences: T1- and T2-weighted images with or without

fat saturation, DCE acquisitions involving multiple post-contrast time points, along with subtraction and maximal intensity projection (MIP) post-processed images (Arian et al., 2022; D'Orsi et al., 2013). The overall breast composition is also described using three main aspects: the amount of fibroglandular tissue, the volume and intensity of background parenchymal enhancement, and presence of implants. Aside from structural information, the DCE acquisition in breast MRI screening also allows radiologists to observe and record functional characteristics of the breast tissue and any possible masses. The BI-RADS atlas also outlines the types of findings that should be noted: foci, masses, non-mass enhancements, and kinetic curve assessment of lesions through which the uptake and washout of the contrast material characterize the lesion type. Background parenchymal enhancement (BPE) is qualitatively assessed as one of four categories: minimal, mild, moderate or marked.

Multiple studies have established the association of BPE with breast cancer risk (Arasu et al., 2019; Dontchos et al., 2015; King et al., 2011; Telegrafo et al., 2016). The exact mechanism by which BPE influences cancer development remains unclear, but it reflects increased leakage of contrast material into the fibroglandular tissue (Kuhl et al., 1997), highlighting areas with physiologically active tissue responsive to hormonal fluctuations, prone to increased inflammation and provide suitable microenvironments for tumorigenesis (Bhatelia et al., 2014). While more research is needed to explicitly investigate the biological mechanism linking BPE and risk of breast cancer development, qualitative BPE associates with cancer risk independently of breast density (Arasu et al., 2019; Dontchos et al., 2015). While qualitative BPE assessment is recorded as part of clinical interpretation of MRI screening, the work of this dissertation explores more specific quantitative BPE metrics computed directly from the MR images.

2.3 Breast cancer risk prediction models

2.3.1 Conventional clinical risk assessment tools

An important aspect of breast cancer screening is the consideration of cancer risk as calculated by risk assessment tools. They are often used to help determine the appropriate screening strategy depending on whether an individual is at average or higher risk. While earlier risk assessment tools assessed risk of carrying a BRCA1 or BRCA2 genetic mutation, such as the Myriad I (Shattuck-Eidens et al., n.d.), Myriad II (Frank et al., 1998), and Penn/Couch models (Couch et al., 1997), I will focus on several current risk models that more directly estimate the risk of cancer. I selected the models discussed in this section based on a literature search (Amir et al., 2003; Brentnall et al., 2015; Cintolo-Gonzalez et al., 2017b; Gail et al., 1989; Jacobi et al., 2009; Ozanne et al., 2013; Ready et al., 2010) and not meant to be an exhaustive list; in Figure 2.1 I will focus on four of the most common models.

Among models that directly predict breast cancer risk, there are the Gail model (Gail et al., 1989), Chen model (Chen et al., 2006), Claus model (Claus et al., 1994),

Barlow model (Barlow et al., 2006), and the Tice/BCSC model (Tice et al., 2008). The Gail model (Gail et al., 1989) is a well-studied model which considers an individual's hormonal exposure history (age at menarche, age at first live birth), breast pathology history (number of prior breast biopsies, history of atypical hyperplasia), as well as genetic (BRCA mutation) and family history of breast cancer. The Chen model (Chen et al., 2006) was an extension from the Gail model at the time of its release by incorporating breast density. The Claus model (Claus et al., 1994) predicts cancer risk by relying on the assumption of autosomal dominant inheritance and without including non-hereditary factors, considering only: age, the number of first- and second-degree relatives with breast cancer, and the age at their diagnoses. The Barlow model (Barlow et al., 2006) requires different risk factors based on menopausal status: premenopausal women need to provide age, breast density, number of first-degree relatives with breast cancer, and previous breast procedures; postmenopausal women need to provide more factors including age, ethnicity, body mass index, age of first live birth, use of hormonal therapy, number of first-degree relatives with breast cancer, breast density, and prior breast procedures. The Tice/Breast Cancer Surveillance Consortium (BCSC) model (Tice et al., 2008) incorporates risk factors similar to the Gail model, with the addition of breast density, but not presence of a genetic BRCA mutation: age, race and ethnicity, age first live birth, personal history of breast biopsy, history of breast cancer in a first- or second-degree relative, and breast density. These models were developed using cohorts designed for different purposes and therefore had different risk factors available.

Among models that predict both breast cancer risk and risk of carrying a genetic mutation, relevant examples include the BRCAPRO model (Mazzola et al., 2015), the BOADICEA model (Lee et al., 2016), and Tyrer-Cuzick model (Tyrer et al., 2004). The BRCAPRO model considers age, race and ethnicity, family history of breast cancer from any degree of relative, the age at their diagnoses, personal or familial ovarian cancer, and the presence of BRCA mutation. The BOADICEA model (Lee et al., 2016) considers age, family history of breast cancer from any degree relative, the age at their diagnoses, presence of male breast cancer, and personal or familial ovarian cancer, and the presence of a series of deleterious mutations: BRCA1, BRCA2, PALB2, CHEK2, and ATM. BRCAPRO and BOADICEA both put a substantial focus on hereditary information and less so on factors such as personal hormonal exposure or breast pathology history. The Tyrer-Cuzick model (Tyrer et al., 2004) is by far the most comprehensive model which includes risk factors from both personal and hereditary aspects: age, race and ethnicity, age at menarche, age at first live birth, age at menopause, parity, hormone therapy, prior biopsy results, presence of BRCA mutation, family history of breast and ovarian cancer, and breast density. This model's ability to consider a wide range of factors in both personal and hereditary aspects is an important reason why we selected it for calculating risk scores in this dissertation work, which we used to predict 5-year cancer risk in Chapter 6 alongside imaging markers. In Figure 2.1, we compare the list of risk factors of four risk models which we found to be most common and relevant, from our literature search (Amir et al., 2003; Brentnall et al., 2015; Cintolo-Gonzalez et al., 2017b; Gail et al., 1989; Jacobi et al., 2009; Ozanne et al., 2013; Ready et al., 2010).

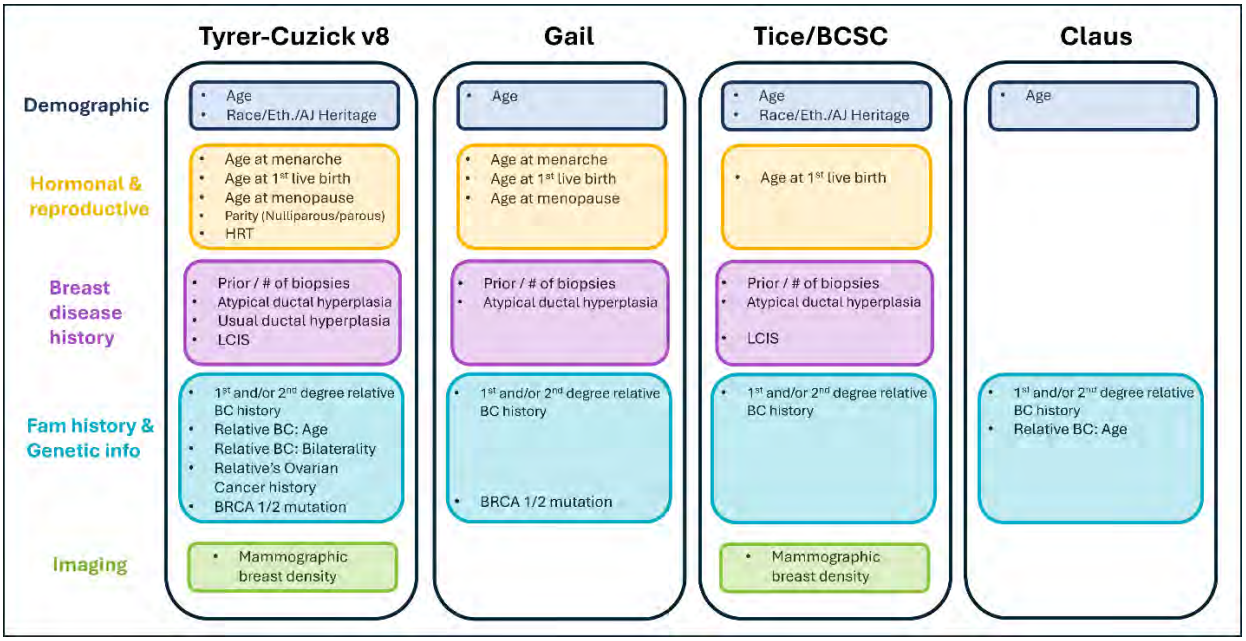

Figure 2.1 A comparison of risk factors used in a few popular clinical risk assessment tools.

The figure illustrates the various combinations of risk factors considered in each risk assessment tool, showing different approaches in predicting cancer risk. For this dissertation work, I selected Tyrer-Cuzick as our primary clinical risk assessment model due to its holistic capture of various risk factors, and also because it is routinely used during breast screening at our institution.

2.3.2 Imaging-based models

Here I will briefly describe two imaging-based risk prediction models. They are both deep-learning based methods which utilize a breast imaging scan as input, and output a quantitative or binary prediction as cancer risk.

Yala et al. developed Mirai, a deep learning model that utilizes mammographic views and clinical risk factor information to predict cancer risk at each year over the next 5 years (Yala et al., 2021). The model was trained on 210,819 exams, validated on 25,644 exams, and tested on 25,855 exams. Of the 25,855 exams in the home institution, 588 examinations were considered “positive” in the sense that they eventually developed pathology-confirmed cancer within 5 years. The model performed well when tested on held-out test sets from their home institution as well as two external institutions, obtaining C-indices of 0.76 to 0.81. Strengths of this model include its large training dataset, externally validated results, and an innovative approach to use clinical risk factors when available and making the model impute the risk factor values when they are unavailable. Main weaknesses include the fact that Mirai was trained using mammograms from a single vendor (Hologic) and so may be limited in its ability to assess mammograms from other vendors, and also that it does not address the interpretability aspect for how it estimated a risk score for a given mammogram.

Arefan et al. developed and compared two mammogram-based deep learning models, a linear discriminant analysis (LDA) model called GoogLeNet-LDA and an end-to-end model called GoogLeNet, using mediolateral oblique (MLO) and craniocaudal (CC) views to predict short-term risk (within 1 to 2 years) of developing cancer (Arefan et al., 2020). Training and testing were performed with 10-fold cross validation on 452 images from 226 patients. The patients involved were a case-control cohort, where 113 cases are patients who later developed breast cancer while the 113 controls are matched controls who did not. The two models were compared across several experiments, with variants such as whole-breast images or dense regions and CC view, MLO view, or combined CC+MLO views. They found that in the highest performing experiment results, the GoogLeNet-LDA model (AUC: 0.73) outperformed the end-to-end GoogLeNet model (AUC: 0.68). Strengths of this study include the use of two different deep learning schemes and the use of case-control cohort design, which matches positive cases to controls, as well as an inclusion of some interpretability in the form of feature map visualization that indicates importance of regions in predicting risk. Weaknesses of this study include the fact that they also used mammograms acquired from a single vendor, and that it is a single-center retrospective study which needs to be validated externally.

These imaging-based models take a different approach from our work; they utilize 2D mammogram images to generate a prediction, while we utilize quantitative markers derived from 3D MRI. Their models rely on mammography and therefore were not trained on MR images; applying these to MRI would require significant re-training using MR images. Using an entire image as input may preserve the rich representation of information, allowing the model to take full advantage of all features present in an image. Mirai does not address interpretability of how each mammogram leads to a specific risk score as result, while Arefan et al. did address that through feature map visualization that identifies regions that were important for risk prediction. Our approach to use specifically extracted quantitative markers enables the selection of features with known characteristics, enabling more focused associations and interpretable results. We believe there are merits to both approaches, each worth exploring and having the potential to make important contributions to the field of breast cancer risk prediction.

Before I begin discussing the study cohort selection for the main results of this dissertation work, I will describe in the next chapter preliminary work in the association of cancer risk with quantitative MR imaging markers, which was accepted as a publication (Surento et al., 2025).

Chapter 3

Preliminary work on quantitative MR markers associated with cancer risk

In this chapter, I present preliminary work done to explore the association between quantitative imaging markers and cancer risk on a cohort of 77 women who did not yet have 5-year cancer outcomes from follow-up. Since the true outcomes were not available, I used low and high Tyrer-Cuzick risk categories as a proxy for cancer risk. Here I investigate the association of quantitative BPE along with other multiparametric MRI-derived metrics with TC risk score categories in a screening population of 77 women with dense breasts; this work was published as a full-length journal article in *Cancers* (Surento et al., 2025).

3.1 Multiparametric MRI markers associated with breast cancer risk in women with dense breasts from a 2019-2022 cohort (n=77)

3.1.1. Methods for studying the association of Multiparametric MRI markers and breast cancer risk (n=77)

Patient Population

This study involved patients enrolled in an IRB-approved research study evaluating the use of diffusion-weighted MRI in screening women with dense breasts. Breast density eligibility was determined based on mammography or MRI (BI-RADS density category C – heterogeneously dense or D – extremely dense). Between November 2019 and May 2022, multiparametric breast MRI examinations were performed on participants for indications of either high-risk screening or problem-solving MR following discovery of a suspicious lesion on routine mammography or MR screening. All participants were confirmed to be negative for cancer, based on biopsy and 1-year follow-up. At our institution, TC risk assessment is routinely performed as part of the clinical workflow during breast screening. TC lifetime risk scores and clinical factors such as age, demographics, and menopausal status were retrospectively extracted from electronic medical records.

Image Acquisition and Markers

Multiparametric breast MRI scans were acquired on a Philips Achieva 3T scanner using a dedicated 16-channel breast coil (Mammotrak, Philips Healthcare, Best, The

Netherlands). Dynamic contrast-enhanced (DCE) MRI involved one pre-contrast and multiple post-contrast T1-weighted scans, with the first post-contrast scan centered at 110 s after contrast injection (0.1 mmol/kg body weight; ProHance, Bracco Diagnostics Inc., Milan, Italy). Diffusion-weighted imaging (DWI) was acquired with b-values 0, 100, 800, and 1200 s/mm^2. The details of the acquisition parameters are in Table 3.1.

Table 3.1. Image acquisition parameters.

Protocol	DCE	DWI
Sequence Type	3D Fast Field Echo	2D single-shot—Echo Planar Imaging
Acquisition plane	axial	axial
Phase encode	R/L	A/P
Field-of-view (mm)	240 × 360	360 × 360
Slice thickness (mm)	1.5	4
Acquisition matrix	480 × 720	200 × 200
In-plane resolution (mm^2)	0.5 × 0.5	1.8 × 1.8
TR (msec)	5.7	3500
TE (msec)	3	75
Flip angle	10	90
Fat suppression	SPAIR	SPAIR
b values, s/mm^2		0, 100, 800, 1200
Gadolinium-based contrast agent	gadoteridol (0.1 mmol/kg body weight)	-
Acquisition time	2 min per phase, (1 pre-contrast, 3 post-contrast phases with k0 at approx. 2, 4, and 6 min after contrast injection) <9 min total	3 min

Qualitative BPE assessments were obtained from imaging reports, where BPE was assessed by radiologists using DCE subtraction images (first post-contrast minus pre-contrast) and maximum intensity projections (MIPs) at the time of clinical interpretation, and categorized as minimal, mild, moderate, or marked in accordance with BI-RADS guidelines (D'Orsi et al., 2013). Background parenchymal diffusion signal (also known as diffusion background signal (Hahn et al., 2016)), defined as the persistent signal from fibroglandular tissue on DWI, was similarly assessed retrospectively by a radiologist using b = 1200 s/mm^2 images (BPDS) and categorized as minimal, mild, moderate, or marked, as previously described (D'Orsi et al., 2013) (Figure 3.1).

Quantitative Imaging Markers

Quantitative imaging markers were calculated using custom software developed in MATLAB R2021b 9.11.0.2358333 Update 7 (Mathworks, Natick, MA, USA). Pre-contrast DCE images underwent preprocessing to correct for patient motion using 2D affine registration and then resampled to isotropic resolution (1 × 1 × 1 mm). Image segmentation was performed using an open-source deep learning-based algorithm (Lew et al., 2024), which reported Dice similarity coefficients of 0.92 for breast, 0.86 for FGT, and 0.65 for blood vessels when compared to manually annotated masks. Our recent evaluations have shown this algorithm to provide accurate segmentations based on radiologist assessment while improving efficiency and reproducibility over semi-automated methods (Kuo et al., n.d.). Using this algorithm, volumetric masks were generated for bilateral whole breasts, FGT, and large, apparent vessels (Figure 3.1). Masks of the whole breast, FGT, and vessels were used to derive volumetric metrics for the analysis (Table 3.2).

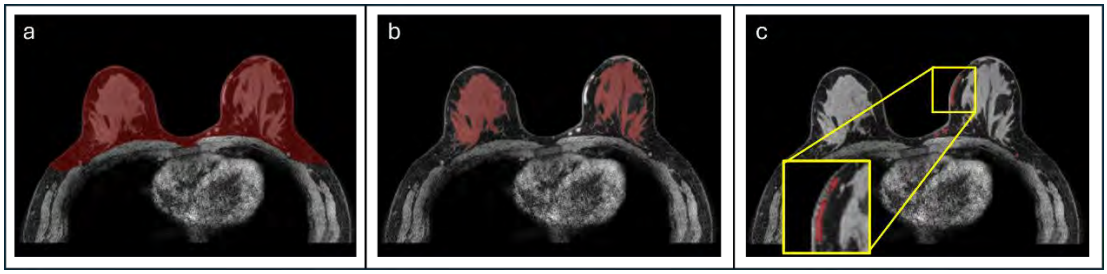

Figure 3.1. Automated volumetric segmentations. Shown are representative masks (red) generated using a deep learning-based segmentation algorithm by Lew et al. (Lew et al., 2024) of (a) bilateral whole breasts, (b) FGT, and (c) larger vessels (as seen in the magnified box, yellow).

Table 3.2. Quantitative BPE and volume measures.

	Derivation Formula
Volume measures	
Breast volume (cm^3)	number of voxels in Breast mask × volume per voxel
FGT volume (cm^3)	# of voxels in FGT mask × volume per voxel
Vessel volume (mm^3)	# of voxels in Vessel mask × volume per voxel
FGT:breast ratio (%)	(FGT volume/Breast volume) × 100%
Vessel:breast ratio (%)	(Vessel volume/Breast volume) × 100%
Vessel:FGT ratio (%)	(Vessel volume/FGT volume) × 100%
BPE measures	
Median PE (%)	median PE of voxels > 10% within FGT mask
BPE volume (cm^3)	# of voxels with PE > 10% within FGT mask × volume per voxel
BPE:breast ratio (%)	(BPE volume/Breast volume) × 100%
BPE:FGT ratio (%)	(BPE volume/FGT volume) × 100%
BPE:vessel ratio	(BPE volume/Vessel volume)
Integrated intensity (cm^3)	BPE volume × mean PE

Percent enhancement (PE) maps were calculated as (S1 − S0)/S0 × 100%, where S0 and S1 are the pre-contrast and first post-contrast DCE-MR images, respectively. A base threshold (PE > 10%) was applied to exclude sporadic noise, only including voxels that exhibited a PE of greater than 10% for analysis. Quantitative BPE calculations were then performed by applying the FGT mask to the PE map (Figure 3.1b), further described in Table 3.2.

Apparent diffusion coefficient (ADC) maps were calculated with b = 0, 800 s/mm^2 using voxel-wise fitting of the monoexponential decay model (Stejskal & Tanner, 1965). Bilateral FGT masks were segmented on b = 0 s/mm^2 images using a custom semi-automated pipeline developed in MATLAB (Mathworks, Natick, MA, USA) based on fuzzy c-means clustering, first identifying the whole-breast area and then the FGT region (Figure 3.2). FGT masks were then applied to ADC maps to calculate summary ADC metrics (median, interquartile range [IQR], skewness).

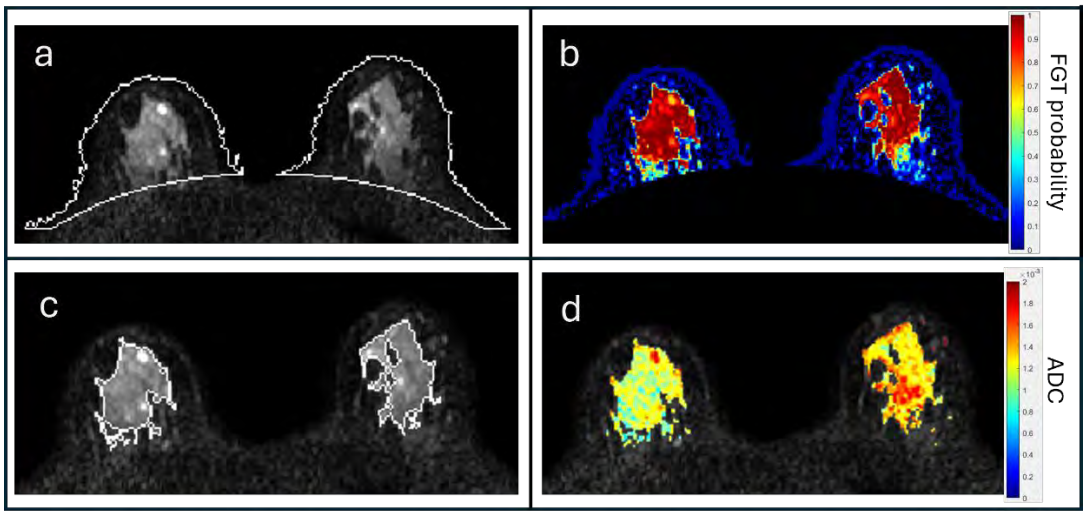

Figure 3.2. Segmentation of FGT on DWI. Shown are outputs from custom segmentation pipeline using *b* = 0 DWI, generating whole-breast masks ((a), white outline on *b* = 0), fuzzy c-means probability map of FGT (b), final FGT mask ((c), white out-line on *b* = 0), and final ADC map of masked FGT region (d).

Statistical Analyses

Four subgroups of imaging markers were defined and evaluated: (1) qualitative markers (BPE and BPDS), (2) tissue volume measures (breast volume, FGT volume, vessel volume, FGT:breast ratio, vessel:breast ratio, vessel:FGT ratio), (3) BPE-related measures (median PE, BPE volume, BPE:breast ratio, BPE:FGT ratio, BPE:vessel ratio, integrated in-tensity), and 4) diffusion-weighted imaging measures (median ADC, ADC IQR, and ADC skewness).

The associations between imaging markers and patient age were evaluated using Spearman's correlation. The TC risk score was used to classify patients into low-risk (≤20%) and high-risk groups (>20%). Discrimination between these groups using individual im-aging markers was assessed using receiver operating characteristic (ROC) curves and summarized using the area under the ROC curve (AUC). DeLong's test was used to compare AUC values between markers.

For the primary analysis, logistic regression analysis was conducted to assess the associations of individual imaging markers with TC group after adjusting for age. Histograms were used to visually assess the distributions of markers across patients. Markers demonstrating right-skewness were log-transformed to reduce asymmetry before including them in the models. Associations were summarized using odds ratios (ORs). We adjusted for age in the modeling since it may affect parenchymal tissue biology and enhancement (Müller-Schimpfle et al., 1997).

Due to the number of imaging markers being tested for statistical significance, it was necessary to adjust the p-values to reduce the risk of false positive findings in the primary analysis. Holm's method was used to adjust p-values from the age-adjusted logistic regression analysis within each imaging marker subgroup (qualitative markers, segmentation-derived volume measures, BPE-related measures, and diffusion-weighted imaging measures). Adjusted p-values (adj. $p < 0.05$) were considered statistically significant in the primary analysis. p-values from other analyses were not adjusted, and statistical significance in these analyses was defined as $p < 0.05$ without adjustment.

Concordance rates between actual TC risk group and predicted TC risk group based on the age-adjusted logistic regression models were further investigated. This analysis was limited to the quantitative markers that had a statistically significant association with TC risk group in the primary analysis and qualitative BPE as a reference. The predicted TC risk group was classified as high risk if the predicted probability of being high risk exceeded the prevalence of the TC high-risk group in the whole cohort (74%) and low risk otherwise.

3.1.2. Results for studying the association of Multiparametric MRI markers and breast cancer risk (n=77)

Patient Characteristics and Follow-Up

Seventy-seven women with dense breasts were identified as eligible and included in the analysis (median age: 45 years, range: 26–71 years), of which 57 (74%) were at high risk of breast cancer. The majority of women were pre- or peri-menopausal (72.7%), had heterogeneously dense breasts (61%), were of White race (82%), and at high risk (74%) (Table 3.3). The proportion of women with extremely dense breasts was greater in the high-risk group (27/57; 47%) compared to the low-risk group (3/20; 15%), aligning with established associations of density with risk. By most recent follow-up at median 26.4

months (range 17.4 to 45.5 months) after MRI, two patients had developed breast cancer: one with TC lifetime risk of 22.9% who was diagnosed with invasive lobular carcinoma at 25.6 months after MRI and another with TC lifetime risk 47.7% who was diagnosed with invasive ductal carcinoma at 27.4 months after MRI.

Table 3.3. Patient characteristics for n=77 cohort.

Variable		Overall Cohort (N = 77)
Age at MRI, years		45 (40, 52)
Menopausal status		
	Pre-menopausal	52 (68%)
	Peri-menopausal	4 (5%)
	Post-menopausal	21 (27%)
Race		
	American Indian/Alaska Native	3 (4%)
	Asian	5 (76%)
	Black or African American	1 (1%)
	White	63 (82%)
	More Than Once Race	1 (1%)
	Unknown/Not Reported	4 (5%)
Breast Density *		
	Heterogeneous	47 (61%)
	Extreme	30 (39%)
Tyrer–Cuzick risk		
	Low (≤20% lifetime risk)	20 (26%)
	High (>20% lifetime risk)	57 (74%)

All data reported in count (%) or median (interquartile range). * Density based on MR or mammography.

Correlation Between Imaging Markers and Age

The imaging markers are summarized in Table 3.4 with their overall distributions and correlations with age. In general, BPE measures (Spearman's rho: −0.29 to −0.47) and FGT-associated volume measures (Spearman's rho: −0.30 to −0.38) were inversely correlated with age (Spearman's rho: −0.14 to −0.47), while vessel-associated volume measures (Spearman's rho: 0.29 to 0.38) and breast volume (Spearman's rho: 0.25) were positively correlated with age. All correlations with age were statistically significant except for median ADC (Spearman's rho: −0.14, $p = 0.23$).

Table 3.4. Spearman's correlations of imaging markers with age.

Variable	Median (IQR)	Spearman's rho vs. Age	*p*-Value
Qualitative metrics			
Qualitative BPE	3 (1, 4)	−0.31	0.006
Qualitative BPDS	2 (1, 3)	−0.39	<0.001
Quantitative metrics			
Volume measures			
Breast volume (cm^3)	1597 (1108, 2501)	0.25	0.029
FGT volume (cm^3)	204 (157, 326)	−0.30	0.007
Vessel volume (mm^3)	13 (12, 14)	0.31	<0.001
FGT:breast ratio (%)	14.2 (7.6, 19.4)	−0.38	<0.001
Vessel:breast ratio (%)	0.5 (0.4, 0.7)	0.29	0.011
Vessel:FGT ratio (%)	4.0 (1.8, 7.2)	0.38	<0.001
BPE measures			
Median PE (%)	28.1 (23.7, 36.1)	−0.40	<0.001
BPE volume (cm^3)	147 (85, 262)	−0.36	<0.001
BPE:breast ratio (%)	7.9 (4.8, 14.0)	−0.47	< 0.001
BPE:FGT ratio (%)	62.0 (54.7, 72.9)	−0.29	0.011
BPE:vessel ratio	13.7 (8.3, 28.5)	−0.47	<0.001
Integrated intensity (cm^3)	49 (26, 105)	−0.43	<0.001
Diffusion-weighted measures			
Median ADC ($\times 10^{-3}$ mm^2/s)	1.61 (1.43, 1.77)	−0.14	0.23
ADC interquartile range ($\times 10^{-3}$ mm^2/s)	0.49 (0.44, 0.57)	0.24	0.040
ADC skewness (mm^2/s)	0.03 (−0.27, 0.32)	−0.26	0.022

IQR = interquartile range.

Association Between Imaging Markers and Tyrer-Cuzick Risk Classifications

Qualitative and quantitative imaging markers stratified by TC risk group are shown in Table 3.5. Qualitative BPE levels and most quantitative volume and BPE measures tended to be higher in the high-risk versus low-risk group, while the opposite trend was observed in vessel-associated measures. After adjusting for age, BPE:breast ratio (OR: 2.71 per 2-fold increase, adj. $p = 0.037$), FGT:breast ratio (OR: 2.59 per 2-fold increase, adj. $p = 0.046$), and BPE:vessel ratio (OR: 4.59 per 2-fold increase, adj. $p = 0.037$) were each individually associated with higher risk based on TC risk score, as illustrated in Figure 3.3. Neither qualitative BPE (OR: 1.79 per 1-level increase, adj. $p = 0.11$) nor qualitative BPDS (OR: 1.31 per 1-level increase, adj. $p = 0.38$) were significantly associated with risk group after adjusting for age. No other volume measure (adj. $p > 0.055$), BPE measure (adj. $p > 0.22$), or diffusion-weighted measure (adj. $p > 0.44$) was significantly associated with risk group (Table 5).

ROC curves for the three significant quantitative imaging markers (AUC: 0.77–79) and qualitative BPE (AUC: 0.69) are shown in Figure 3.4. While these three quantitative markers provided numerically greater discrimination between low- and high-risk groups than qualitative BPE (AUC: 0.69), the AUC values were not statistically significantly different ($p > 0.22$ for each by DeLong's test).

Table 3.5. Association of each imaging marker with cancer risk group.

Variable	Low Risk (n = 20)	AUC (95% CI)	Adjusted OR † (95% CI)	Adjusted *p*-Value ‡
Qualitative markers				
Qualitative BPE	2 (1, 3)	0.69 (0.57–0.82)	1.79 (1.00–3.39)	0.11
Qualitative BPDS	2 (1, 2)	0.63 (0.50–0.76)	1.31 (0.72–2.50)	0.38
Quantitative markers				
Volume measures				
Breast volume (cm^3)	1918 (1355, 2833)	0.62 (0.49–0.76)	0.74 (0.42–1.27)	0.28
FGT volume (cm^3)	178 (133, 204)	0.70 (0.58–0.83)	1.83 (1.01–3.52)	0.22
Vessel volume (mm^3)	11 (8, 15)	0.66 (0.53–0.79)	0.53 (0.24–1.02)	0.24
FGT:breast ratio (%)	7.4 (6.8, 13.5)	0.77 (0.66–0.88)	2.59 (1.34–5.56)	**0.046**
Vessel:breast ratio (%)	0.6 (0.5, 0.7)	0.68 (0.55–0.80)	0.44 (0.15–0.97)	0.24
Vessel:FGT ratio (%)	7.1 (4.6, 9.3)	0.76 (0.65–0.87)	0.28 (0.09–0.67)	0.055
BPE measures				
Median PE (%)	25.9 (23.7, 38.0)	0.55 (0.39–0.71)	0.91 (0.52–1.64)	>0.9
BPE volume (cm^3)	97 (64, 133)	0.72 (0.60–0.85)	1.81 (1.00–3.46)	0.22
BPE:breast ratio (%)	4.8 (3.5, 6.6)	0.78 (0.67–0.89)	2.71 (1.37–5.98)	**0.037**
BPE:FGT ratio (%)	59.3 (53.7, 68.0)	0.59 (0.45–0.74)	1.19 (0.68–2.06)	>0.9
BPE:vessel ratio	8.3 (5.8, 12.1)	0.79 (0.68–0.90)	4.59 (1.75–15.79)	**0.037**
Integrated intensity (cm^3)	36 (20, 49)	0.66 (0.52–0.80)	0.69 (0.22–1.15)	0.66
Diffusion-weighted measures				
Median ADC ($\times 10^{-3}$ mm^2/s)	1.56 (1.38, 1.66)	0.61 (0.47–0.76)	1.34 (0.79–2.37)	0.57
ADC interquartile range ($\times 10^{-3}$ mm^2/s)	0.48 (0.44, 0.56)	0.51 (0.36–0.66)	1.32 (0.73–2.58)	0.57
ADC skewness (mm^2/s)	0.12 (−0.07, 0.33)	0.59 (0.45–0.73)	0.66 (0.37–1.15)	0.44

AUC = area under the curve; CI = confidence interval; OR = odds ratio; * Values are bilateral and given as median (interquartile range).† Odds ratios are adjusted for age and expressed as the change per 1-level change (qualitative BPE and BPDS) or per 2-fold

change (all quantitative markers).‡ *p*-values were adjusted using Holm's method within subgroups of markers as indicated by the table subheadings. Bold indicates statistical significance after adjustment (adjusted $p < 0.05$).

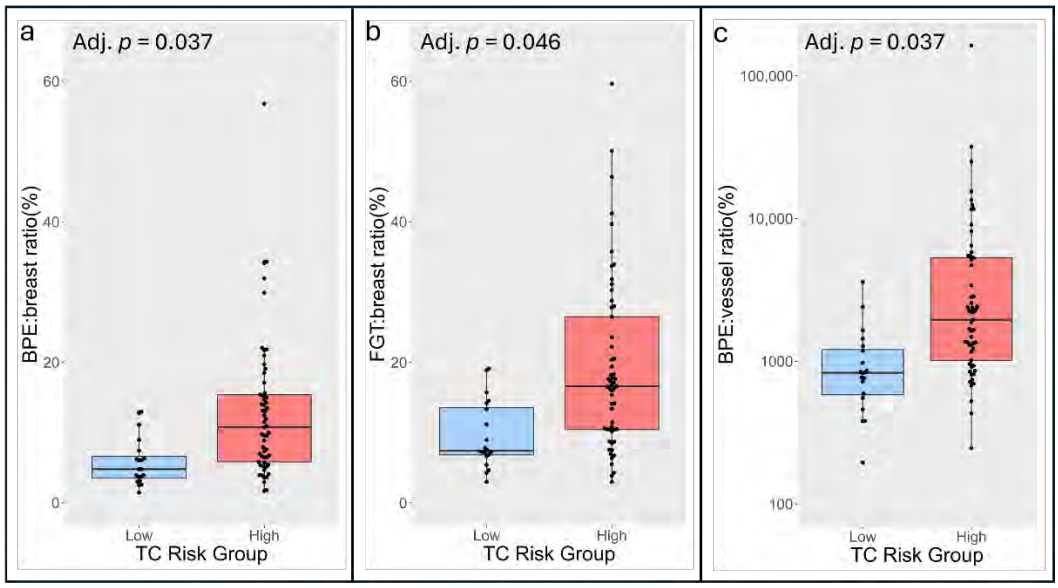

Figure 3.3. Risk-stratified distributions of select imaging markers. Age-adjusted logistic regression analysis demonstrated significant differentiation between low- and high-risk groups (adjusted *p*-values < 0.05). Imaging markers shown are BPE:breast ratio (a), FGT:breast ratio (b), and BPE:vessel ratio((c), shown on $\log_{10}$ scale).

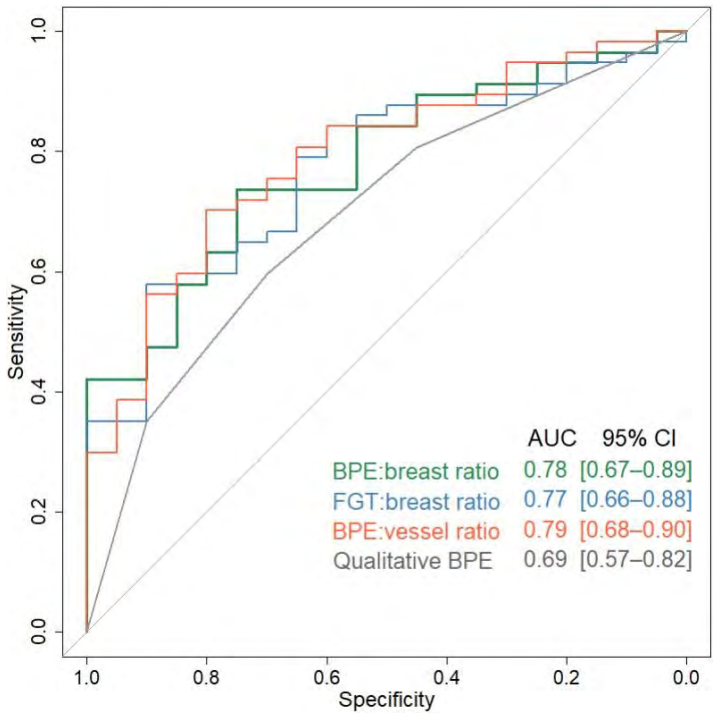

Figure 3.4. Diagnostic performance of imaging markers. ROC curves for discriminating between low- and high-risk groups for quantitative imaging markers that were statistically significant in the age-adjusted primary analysis (Table 3.5) as well as qualitative BPE for reference.

Concordance Between Imaging Markers and Tyrer-Cuzick Risk Classifications

We examined the concordance rates between actual and predicted TC risk group based on the age-adjusted logistic regression models to better understand agreement between MRI markers and Tyrer–Cuzick risk classifications, focusing on qualitative BPE and the statistically significant quantitative markers (Table 3.5). Across the markers, overall concordance rates reached up to 70% and were similar within low- (65–75%) and high-risk groups (67–70%), representing moderate concordance overall. Figure 3.5 presents different examples of concordance between PE and TC risk alongside discordance between PE and TC risk.

Table 3.6. Concordance rates of select quantitative markers and qualitative BPE.

Marker	Overall Cohort (n = 77)	Low Risk (n = 20)
BPE:breast ratio	52 (67%)	14 (70%)
FGT:breast ratio	54 (70%)	14 (70%)
BPE:vessel ratio	53 (69%)	15 (75%)
Qualitative BPE	52 (67%)	13 (65%)

Values are number (%) of cases where the actual and predicted risk groups are equal.

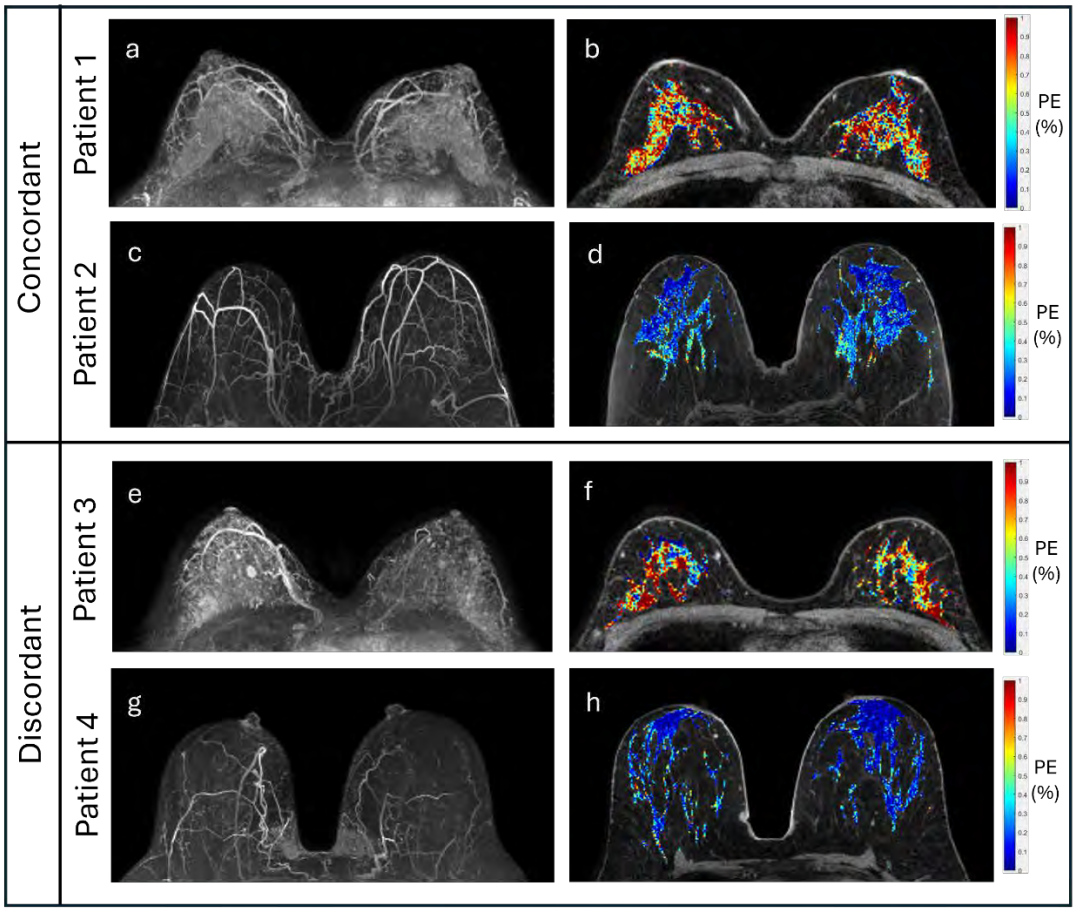

Figure 3.5. Examples of concordant and discordant cases. Subtraction MIPs (left column) and quantitative PE maps (right column) for several patients with concordant and discordant risk categorizations based on imaging markers and TC risk score. Patients 1 and 2 are examples of concordant cases. Patient 1 (**a**,**b**) is a 44-year-old woman with high TC risk (30%) due to past history of lobular carcinoma in situ and atypical lobular hyperplasia, exhibiting marked BPE and BPE:breast ratio of 12.5%. Patient 2 (**c**,**d**) is a 69-year-old woman with low TC risk (16%), exhibiting mild BPE and BPE:breast ratio of 3.1%. Patients 3 and 4 are examples of discordant cases. Patient 3 (**e**,**f**) is a 44-year-old woman with low TC risk (18%), exhibiting moderate BPE and BPE:breast ratio of 13.0%. Patient 4 (**g**,**h**) is a 60-year-old woman with high TC risk (27%) due to family history of breast cancer, exhibiting minimal BPE and BPE:breast ratio of 3.6%.

3.1.3. Discussion for studying the association of Multiparametric MRI markers and breast cancer risk (n=77)

Our study explored a wide range of breast MRI-derived parameters and found that several quantitative MRI markers incorporating BPE and anatomical volumes could be valuable tools to estimate breast cancer risk. Specifically, we found BPE:breast, FGT:breast, and BPE:vessel volume ratios differentiated between low and high Tyrer–Cuzick risk groups (AUCs: 0.77–0.79) and retained significant association with risk after adjusting for age and multiple comparisons. Qualitative BPE, on the other hand, demonstrated a weaker association with risk (AUC: 0.69). Overall, our work validates the findings from prior studies identifying BPE as a marker of breast cancer risk (Arasu et al., 2019; Dontchos et al., 2015; Hu et al., 2017, 2021; Kim et al., 2015; King et al., 2011; Lam et al., 2019; Liao et al., 2020; Niell et al., 2021; Pujara et al., 2018; Telegrafo et al., 2016; Wang et al., 2023) and adds new evidence relating to the potential of other alternate quantitative imaging markers that may be associated with risk. Importantly, these quantitative imaging measures may serve as objective biomarkers that could be incorporated into clinical risk assessment algorithms to improve tailoring of screening and preventative interventions based on a more precise estimate of an individual woman's risk of developing future breast cancer.

Our study further confirms BPE to be associated with a woman's breast cancer risk. While the exact mechanism by which BPE is associated with breast cancer development remains unclear, it is thought that BPE is a marker of physiologically active tissue responsive to hormonal fluctuations. Higher BPE is caused by an increase in permeability of vessels and uptake of contrast by the fibroglandular tissue in response to the effects of estrogen (Kuhl et al., 1997; Zeppa, 1969). These highlighted areas of physiologically active breast tissue are perhaps prone to increased inflammation and linked to favorable microenvironments for tumorigenesis (Bhatelia et al., 2014), therefore elevating cancer risk. The association between increased BPE and cancer risk is reflected in our finding of women who have a higher BPE:breast ratio being nearly three times more likely to be in the high cancer risk group by Tyrer–Cuzick (OR: 2.71, adj. p = 0.037). Our observations are consistent with those found by others investigating quantitative measures of BPE, where relative volumes of BPE differentiated women who subsequently developed cancer versus those who did not (Lam et al., 2019; Niell et al., 2021).

A novel finding of our study was the strong predictive value of the BPE:vessel ratio for breast cancer risk. The critical role of angiogenesis in breast cancer is well documented (Castañeda-Gill & Vishwanatha, 2016; Fakhrejahani & Toi, 2014; Kristensen et al., 2014; Madu et al., 2020; Vartanian & Weidner, n.d.; Weidner et al., 1991), and some studies have identified breast vascularity markers associated with cancer prognosis (Kim et al., 2015) and recurrence (Choi et al., 2016). However, to the best of our knowledge, specific vascular features in the breast preceding primary cancer development have not yet been investigated for their potential link to breast cancer risk. Studies have shown that higher BPE correlates with elevated levels of angiogenesis

markers, such as vascular endothelial growth factor (VEGF) (Kim et al., 2015; Sung et al., 2018). Women with fatty breasts have also been reported to exhibit increased breast vessel density (Fuller et al., 2019), which may also be linked to higher body mass index (BMI) and increased age (Fuller et al., 2019; Jesinger et al., 2011). This relationship may be reflected in our observation that patients of the low-risk group, who had lower breast densities compared to the high-risk cohort (15% extremely dense versus 47%, respectively), also tended to have higher vessel volume measures (though not statistically significant). Taken together with prior work, these findings emphasize the potential relevance of vascular changes, not only in tumorigenesis but also in anatomic and structural predispositions that may help stratify breast cancer risk.

Given that BPE is known to be associated with breast cancer risk (Arasu et al., 2019; Niell et al., 2021; Telegrafo et al., 2016; Wang et al., 2023), we were interested to examine the association of BPE and other MRI markers with TC-based risk classification. We used TC as a surrogate for breast cancer risk since it is among the most comprehensive risk assessment models considering a broad selection of risk factors such as family history, hormonal profile, and biopsy results (Cintolo-Gonzalez et al., 2017a), and it is used clinically for risk assessment at our institution. However, TC does not incorporate imaging markers beyond mammographic density, thus motivating our work to examine its association with BPE and other MRI markers. We found overall good agreement between risk classifications based on MRI marker predictions and TC risk scores, reaching up to 70% concordance. However, lack of concordance in 30% or more of cases also suggests they provide different and potentially complementary information that could be useful in combination to improve accuracy for risk prediction, which is important to explore in future work in study cohorts with known cancer outcomes.

There were several limitations to our study. First, our study cohort included women who were mostly at high risk of breast cancer as well as some who had suspicious findings on mammogram or ultrasound. Also, our sample size was relatively small with a higher proportion of high-risk individuals, which may limit the generalizability of our findings to a broader population of average risk women. Our study defined cancer risk based on Tyrer–Cuzick lifetime risk score as we lacked complete 5-year follow-up and cancer outcomes in our patient cohort. Future larger studies with longer follow-up periods and pathology-confirmed cancer outcomes are needed to confirm associations of imaging markers with cancer risk. Such a study would also potentially demonstrate the incremental value provided by imaging markers and how best to combine them with clinical risk assessments (e.g., TC risk scores) to maximize predictive accuracy. Qualitative BPE was manually assessed by variable radiologists at the time of clinical interpretation; therefore, its performance as a marker may have been limited by inter-observer variability compared to BPE assessments by a single reader. Lastly, the vessel segmentation masks captured visually apparent vessels and may have excluded smaller vessels contained within FGT regions. Therefore, the resulting vessel segmentation mask should not be considered a comprehensive vascular map.

3.1.4. Conclusion for studying the association of Multiparametric MRI markers and breast cancer risk (n=77)

Our study investigated the potential value of multiparametric MRI markers for stratifying breast cancer risk in women with dense breasts. The findings demonstrated that several quantitative MRI markers were significantly associated with risk based on TC risk classifications. Quantitative BPE measurements distinguished women at low risk vs. high risk for breast cancer with higher accuracy than radiologists' qualitative BPE assessments, with the additional advantage of offering more precise and less subjective evaluations. Beyond BPE, we also identified new markers relating to breast vessel volumes that were associated with risk, which may be worth exploring further. As our next step, we plan to perform a larger retrospective study in patients with known cancer outcomes to confirm the predictive value of these MRI markers and assess their incremental value to improve performance of conventional clinical risk assessment models.

The findings described in this chapter led us to expand the work into a larger cohort with 5-year follow-up outcomes that were available. Since I identified several quantitative imaging markers that were associated with TC risk group, the next step was to compare them with actual cancer outcomes in our study cohort. Therefore, in the next chapter, I will describe the approach we used to select our study cohort, and the building of a data repository to store the clinical risk factors relevant in calculating TC risk scores for patients in our cohort.

Chapter 4

Building a data repository for the study cohort

4.1 Cohort description: high-risk screening population

4.2 A data repository for the study cohort

4.2.1 Clinical risk factors to collect

4.2.2 Data Completeness and Summary

In order to build and assess risk prediction models, we need a clean data set that includes both clinical data and imaging information from our study cohort. Here, I describe the work carried out to build this dataset, as well as the cohort selection and sampling approach we used.

As a centralized repository of data for the study cohort, the database enables us to store all pertinent information extracted from electronic medical records, and to keep track of collection progress. Section 4.1 describes the cohort selection and sampling approach used to arrive at our study cohort of 234 individuals. We performed case-cohort sampling in order to arrive at a workable sample size to extract quantitative imaging markers and perform risk factor lookup. Later in this chapter, I will describe the case-cohort sampling approach in greater detail. Section 4.2 details the organization of the REDcap data repository and summarizes what was found to be available or missing from the collection effort.

Ultimately, the goal of collecting clinical risk factors was to calculate Tyrer-Cuzick (v8) risk scores using the IBIS Breast Cancer Risk Evaluation Tool (CharterHouse Square, London, United Kingdom). These risk scores, when combined with imaging markers from Chapter 5, enabled us to perform multimodal modeling of cancer risk in Chapter 6. While Tyrer-Cuzick (TC) risk scores were generated at the time of screening and available in the medical records for a subset of the patients, they were not usable for our purposes due to the lack of recording of individual risk factors that were used for input. In order to harmonize the risk predictions across patients, all TC risk scores used in our work were re-calculated using the clinical risk factors extracted from electronic medical records, enabling us to appraise predictive performance while taking into consideration the extent of data missingness. In addition, collecting risk factors enables us to characterize the patients in our study cohort as well as perform subanalyses based on these specific characteristics, which we explore in Chapter 6.

In addition to describing the organization of this data repository, this chapter may also potentially provide a reference guide for future risk prediction studies involving multimodal combinations of variables.

4.1 Cohort description: high-risk screening population

As outlined in Chapter 2, our work focuses on high-risk women instead of women at average risk of developing cancer. Individuals are at elevated risk of developing breast cancer due to the presence of family history and/or personal genetic mutation; being at high risk makes them eligible for screening with MRI in addition to annual mammography. This is in contrast to average-risk women who receive only mammographic screening.

We outline the process to arrive at our study cohort as follows:

Step 1: Acquire a list of all breast MRI scans from 2005 to 2021, which are accessible to us.

Step 2: Exclude patients known to have personal history of breast cancer according to CSS tumor registry and UW medical records.

Step 3: Exclude patients whose MRI is within 1 year of a subsequent cancer diagnosis, to rule out concurrent cancers.

Step 4: Case-cohort sampling on this pool of women: randomly select 15% of the pool of women from step 3 and include any positive cases not captured in the 15% sampling.

At the end of Step 3, we have identified a pool of patients without prior breast cancer or within 1 year of MRI. Positive cases are defined as those who developed cancer at least one year after and within five years since the MRI screening, and negative controls are those who do not have cancer diagnosis within that period according to follow-up information. Figure 4.1 illustrates the selection process from Steps 1-3.

The risk prediction in this work refers to the model output of a binary outcome of “yes” or “no” cancer, specifically within the period of 5 years since the time of screening MRI. The predicted binary outcome is compared against a true outcome of presence or absence of cancer diagnosis according to follow-up. Time intervals of 5- and 10-year risks are commonly used by breast cancer risk assessment tools because they allow for an actionable timeline for introducing clinical interventions or screening strategies (Cintolo-Gonzalez et al., 2017a; Gail, 2015). We choose to estimate 5-year risk in our work due to a threefold motivation: 1) finding a cohort with 5-year follow-up with cancer outcomes enables a larger cohort size because it is easier than finding a cohort that has 10 complete years of follow-up, 2) improves ability to compare performance with other common risk assessment tools, and 3) minimizes the likelihood of competing mortality due to other health conditions.

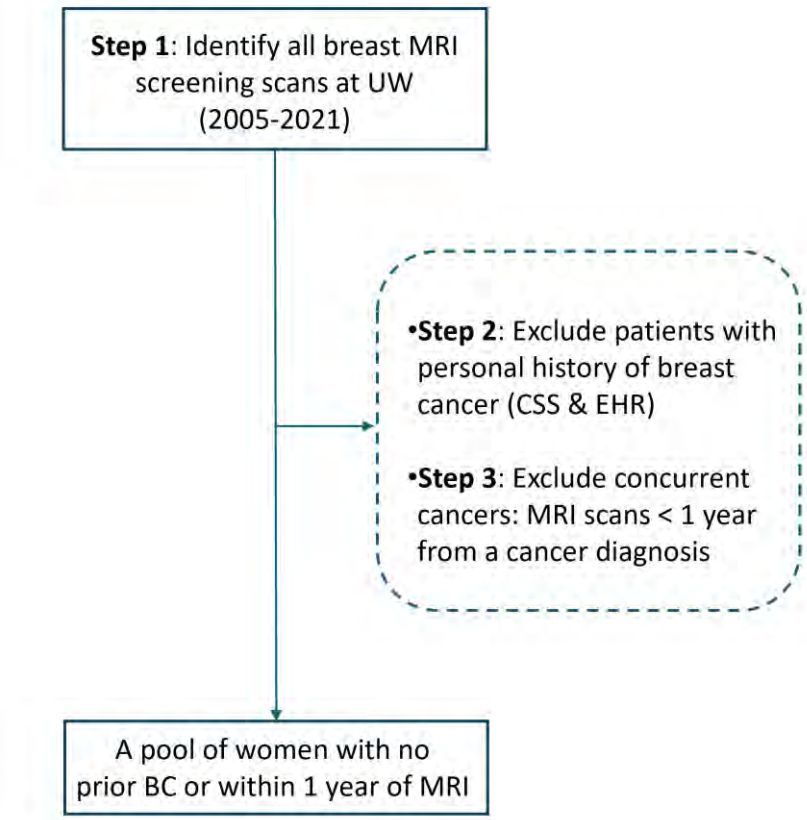

Figure 4.1 Identifying a pool of women with no prior BC or within 1 year of MRI through exclusion criteria (Steps 1-3) in preparation for case-cohort sampling.

Using this pool of patients identified at the end of Step 3, we performed a case-cohort sampling (Step 4) that enables us to leverage findings from a subset to offer insights into the larger screening population. Another benefit to using a subset is to improve feasibility of deriving quantitative imaging markers and collecting clinical risk factors, which would demand far more time and effort if done in the entire pool of patients. Case-cohort sampling is an efficient approach to study cancers and rare diseases, including all incidence cases as positives, and negatives sampled from the initial pool regardless of future disease outcome. This allows us to preserve all positive cases to be used in the final study cohort, improving the feasibility of collecting complete covariate data just for the study cohort instead of the larger pool. The final sample is then enriched with positives: having a relative proportion of positive-to-negative cases that is higher than the observed incidence rate in the general population and regular cancer yield of 2.2 – 3.5% from MRI screening (Lehman et al., 2007; Lehman & Smith, 2009).

Based on sample size projections in preparation for our project, assuming that each quantitative BPE marker associates with new breast cancer by having a true hazard ratio of 1.5 per 1-SD increase in the marker, a sampling fraction of 15% would still grant us 80% power to statistically confirm that association (Cai & Zeng, 2004, 2007). Additionally, applying a 15% sampling fraction also increases the feasibility to produce a

well-curated study cohort with high-quality quantitative markers and manually collected clinical risk factors. A larger sampling percentage would benefit statistical power but also increase the amount of time and effort involved in collecting clinical risk factors and generating quantitative imaging markers.

Figure 4.2 illustrates the case-cohort sampling approach of Step 4 as follows. We begin with a randomly selected 15% sample, which includes mostly negatives and perhaps some positives. Next, we included all women who eventually developed breast cancer later who were not captured in the 15% sampling. This produces 52 positive cases and 182 negatives, totaling in 234 women as the final study cohort. The relative proportion of positive cases is now approximately 22%. This positive-enriched cohort allows our predictive model to have enough positive cases to learn from and result in better predictive performance than if there were fewer positive cases.

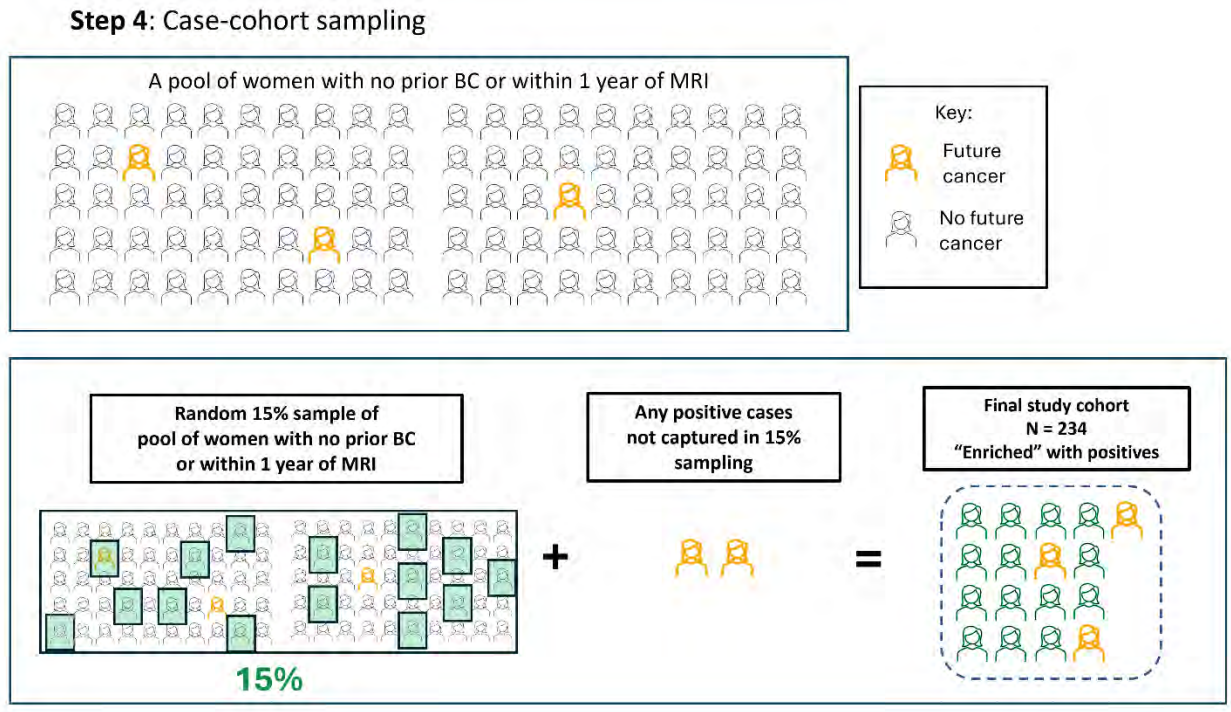

Figure 4.2 Case-cohort sampling (Step 4) approach to obtain a final study cohort of 234 women.

4.2 A data repository for the study cohort

We built a database to store information on the study cohort's 234 women, serving as a repository of clinical factors (age, menopausal status, etc), and imaging information (study accession numbers, date, specific imaging interpretations from radiology reports). Specifically, for positive cases, we also collect histology information: invasive carcinoma or ductal carcinoma in situ (DCIS). The information stored in this database represents a central repository of data from multiple sources, which includes Pathology reports, MRI reports, progress notes, tumor registry, each varying in the extent of data structuredness.

4.2.1 Clinical risk factors to collect

Table 4.1 summarizes the information we collected for the study cohort, as organized in the REDcap data repository we used as a storage platform. The Tyrer-Cuzick assessment tool can take in a large number of risk factors to calculate risk for a single patient. This can range from highly relevant personal risk factors such as biopsy history to cursory factors detailing extended family relatives, such as the number of paternal nieces who had ovarian cancer and the age at which they were diagnosed. Availability of these risk factors may not be uniform across all patients and years of medical recordkeeping. Since we had to extract risk factors via manual lookup from the medical records, we chose to prioritize a group of core risk factors that are commonly used by other risk assessment tools.

They included:

- age at MRI scan,
- age at menarche,
- age at menopause,
- menopausal status at the time of MRI scan,
- BRCA test result, mammographic breast density,
- prior result of biopsy procedure
- family history of breast cancer
- family history of ovarian cancer.

While Tyrer-Cuzick is the main clinical risk model used in our work, we aimed to support future use of other clinical risk assessment tools. Because cancer risk models rely on different clinical factors (Chapter 2), we selected core risk factors that balance overlap across models with data collection feasibility. This approach allows the same dataset to be used to calculate risk scores from multiple assessment tools.

Table 4.1 Patient clinical and imaging information stored in the data repository

Data field group	Data field name	Data type	Imaging or Clinical	TC model input (Y/N)
Demographics				
	Age at MRI scan	Numeric	Clinical	Y
	Age at menarche	Numeric	Clinical	Y
	Age at onset of menopause	Numeric	Clinical	Y
	Menopausal status at MRI scan	Categorical: • Premenopausal • Perimenopausal • Postmenopausal • Unknown	Clinical	Y
	BRCA test result	Categorical: • BRCA 1 • BRCA 2 • Negative • Unknown	Clinical	Y
	Prior result of biopsy procedure	Categorical: • Lobular carcinoma in situ (LCIS) • Atypical hyperplasia • Hyperplasia (not atypia) • Prior biopsy result unknown • No prior biopsy/no proliferative disease	Clinical	Y
	Family history of breast cancer: Relative type	Categorical: • Mother • Sister • Half-sister • Daughter • Grandmother	Clinical	Y

		• Niece • Etc Numeric (age at diagnosis)		
	Family history of breast cancer: Degree of relative	Categorical: • First degree only • Second degree only • At least two first degree • At least one first degree & at least one second degree • Third degree only • No family history • Unknown Numeric (age at diagnosis)	Clinical	N
	Family history of ovarian cancer	Categorical (relative type): • Mother • Sister • Half-sister • Daughter • Grandmother • Niece • Etc Numeric (age at diagnosis)	Clinical	Y
Diagnosis (for positive cases only)				
	Date of diagnosis	Date (mm/dd/yyyy)	Clinical	N
	Histology	Categorical: • Invasive ductal carcinoma • Invasive lobular carcinoma	Clinical	N

		• Ductal carcinoma in situ (DCIS)		
Imaging				
	Mammographic breast density	Categorical: • Fatty • Scattered • Heterogeneously dense • Extremely dense	Imaging	Y
	MRI Accession Number	Numeric/Integer	Imaging	N
	MRI Date	Date (mm/dd/yyyy)	Imaging	N
	Qualitative background parenchymal enhancement	Categorical: • Minimal • Mild • Moderate • Marked	Imaging	N

4.2.2 Data completeness and summary

Many of the clinical risk factors collected did not exist as structured data in the EHR and needed to be derived or manually discerned from free text. This involved reading relevant Progress Notes for all 234 individuals. For instance, to collect our risk factor of "Family history of breast cancer: Degree of relative", it was necessary to find a progress note which mentions that, read prose text describing the individual's family history, and translate that into the corresponding category such as "first degree only" or "at least two first degree", etc., to be recorded in our data repository. Additionally, we took note of the date when the risk factor is recorded into the electronic medical records wherever possible, such that we only collect risk factors that were available at the time of the MR scan and leave it as "unknown" if it was ascertained after the date of MR scan. For instance, if there is a record of biopsy performed after the MR date, we would not collect that biopsy result because it would not be relevant at the same time as MRI-based risk assessment.

Due to the unstructured nature of many of these clinical risk factors, they may not always be found for every individual, leading to varying levels of completeness across the

cohort. Age at MRI scan was an exception; it could be derived from the date of MRI and therefore was available for everyone in our study cohort. It therefore became a relevant aspect of our work to assess the extent of missingness for these clinical risk factors. Next, we present a selection of clinical risk factors and their completeness via pie charts. In the tables preceding each pie chart, we report breakdowns of cases and controls. The Appendix Section includes two more completeness pie charts: age at menarche and family history of ovarian cancer.

Age at menopause

Age at menopause was also recorded in free text and may not have been readily available for a number of patients in the study cohort. When clinical notes indicate postmenopausal with no date of menopause given, the date was inferred as 12 months after the last recorded date of menstrual period.

	Total (n=234)	Cases (n=52)	Ctrl (n=182)
Found/Derived	187(80%)	39 (75%)	148 (81%)
Missing/Unknown	47 (20%)	13 (25%)	34 (19%)

*All percentages calculated per column and rounded to the nearest whole number

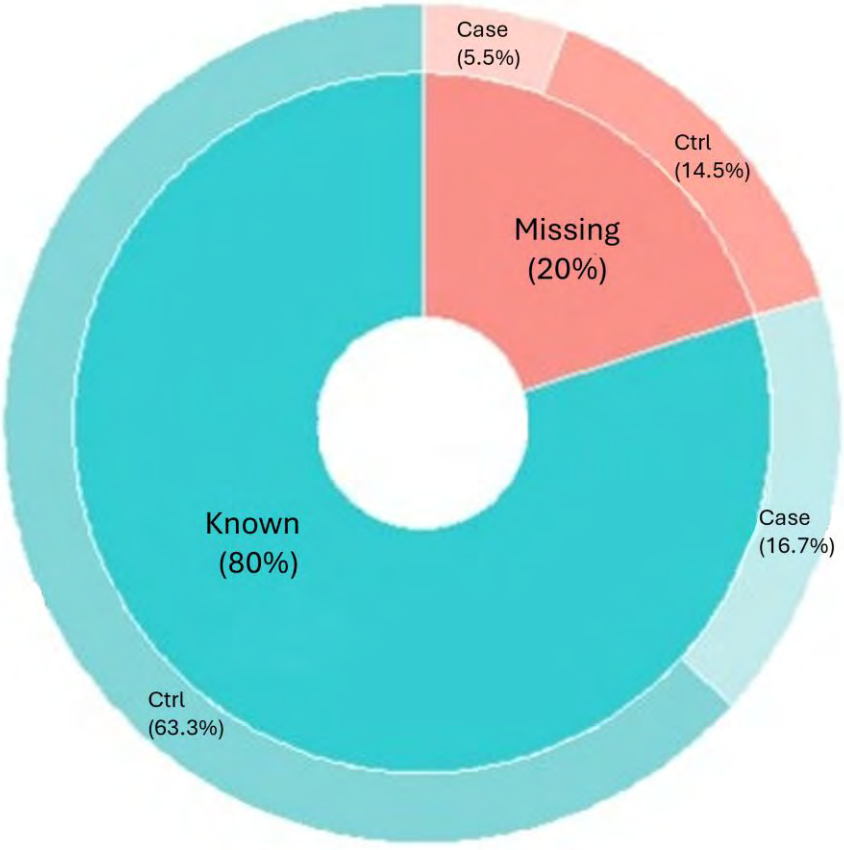

Figure 4.3 Pie chart for distribution of age at menopause, along with breakdown for positives (Case) and negatives (Ctrl).

Menopausal status at MRI scan

Similar to age at menopause, this risk factor sometimes needed to be derived from a date of onset of menopause by adding 12 months to the date of last recorded menstrual period. Close to 60% of the women in the study cohort are premenopausal, about 37% are postmenopausal, 1% are perimenopausal, and 3% are unknown. It is worth noting that we have observed that women are more commonly categorized as premenopausal or postmenopausal.

	Total (n=234)	Cases (n=52)	Ctrl (n=182)
Premenopausal	138 (59%)	35 (67%)	103 (57%)
Perimenopausal	3 (1%)	-	3 (2%)
Postmenopausal	86 (37%)	17 (33%)	69 (37%)
Missing/Unknown	7 (3%)	-	7 (4%)

*All percentages calculated per column and rounded to the nearest whole number

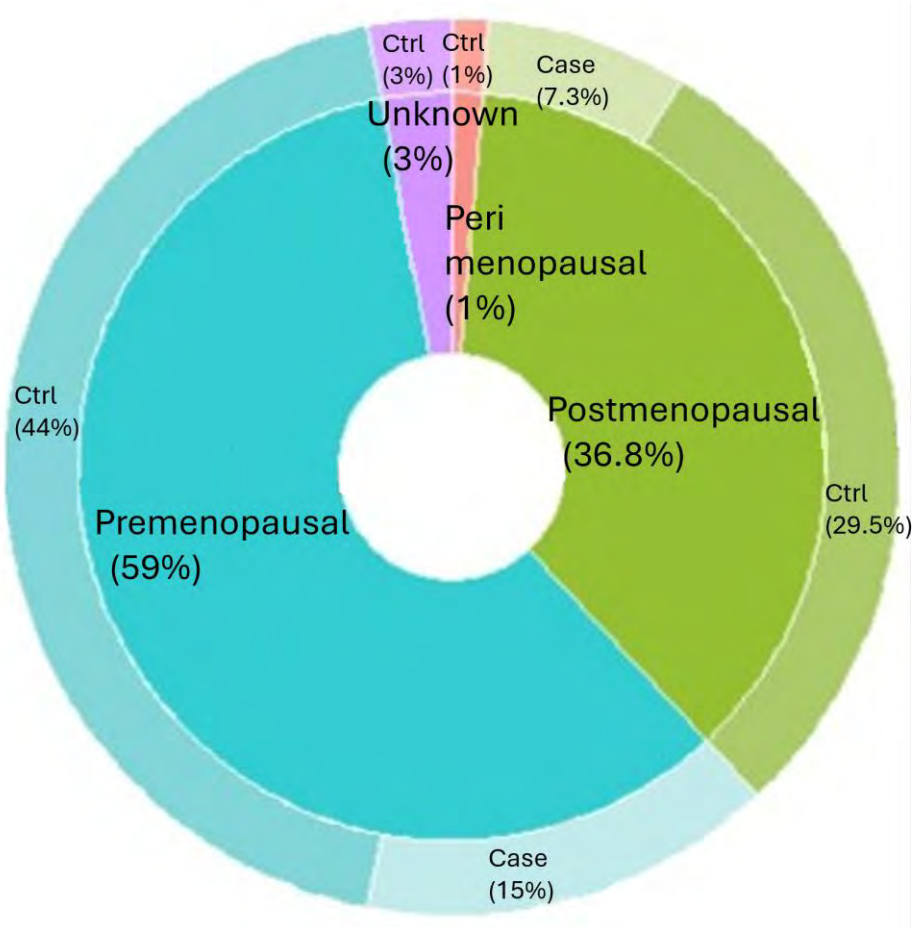

Figure 4.4 Pie chart for distribution of menopausal status at MRI scan, along with breakdown for positives (Case) and negatives (Ctrl).

BRCA Test Results

We found that BRCA testing results were not performed for close to three quarters of our study cohort. Common reasons for no BRCA test performed include omittance due to non-coverage by insurance, or that several first-degree relatives have tested negative and therefore the individual was deemed unlikely to carry a BRCA mutation herself. Results of BRCA testing greater than 1 year after the MRI date were ignored (considered "Unknown" for our study) because it would be unfair to compare the predictive performance of MRI markers to TC risk scores generated using BRCA test results available long after imaging. Another benefit of finding BRCA test results for our study cohort is the ability to exclude BRCA-positive individuals for subset analysis later in Chapter 6. Since a positive BRCA test result mostly outweighs other risk factors, excluding them from analysis enables us to focus on those who would benefit the most from risk modeling. Moreover, the understanding that tumorigenesis originating from BRCA mutation is likely different from those who do not have BRCA mutation further warrants separate evaluation of model performance (Grubstein et al., 2018).

Of 63 patients with testing results in our cohort, 29% tested negative for BRCA mutations and 71% tested positive.

	Total (n=234)	Cases (n=52)	Ctrl (n=182)
BRCA 1	28 (12%)	10 (19%)	18 (10%)
BRCA 2	17 (7%)	5 (10%)	12 (6%)
Negative	18 (8%)	4 (8%)	14 (8%)
Missing/Unknown	171 (73%)	33 (63%)	138 (76%)

*All percentages calculated per column and rounded to the nearest whole number

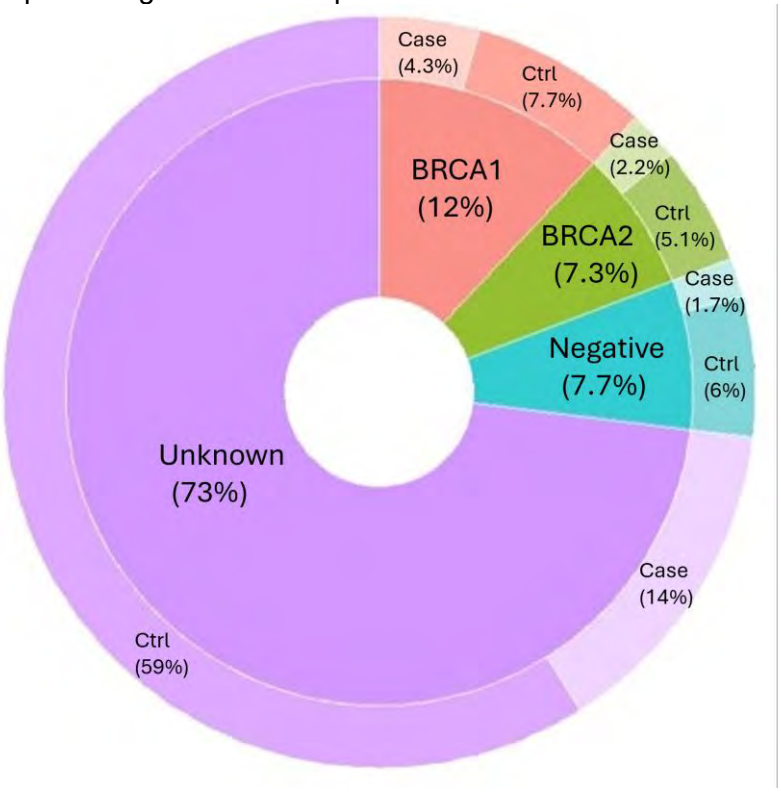

Figure 4.5
Pie chart for distribution of BRCA test results, along with breakdown for positives (Case) and negatives (Ctrl).

Prior Biopsy Test Results

About 85% of the patients in our study cohort are classified as “No prior biopsy/No proliferative disease”. Among the 85%, there are also some patients with missing biopsy results, but are put into this category because the Tyrer-Cuzick version 8 model does not have an option for missing or unknown. Patients with missing biopsy results are categorized as no prior biopsy for TC calculation. The remaining is distributed amongst various high-risk lesions, each approximately numbering in single digit percentages.

	Total (n=234)	Cases (n=52)	Ctrl (n=182)
No prior biopsy/No proliferative disease	200 (85%)	43 (82%)	157 (86%)
Prior biopsy, result unknown	4 (<2%)	1 (2%)	3 (2%)
Hyperplasia (not atypia)	6 (3%)	2 (4%)	4 (2%)
Atypical hyperplasia	20 (9%)	4 (8%)	16 (9%)
Lobular carcinoma in situ (LCIS)	4 (<2%)	2 (4%)	2 (1%)

*All percentages calculated per column and rounded to the nearest whole number

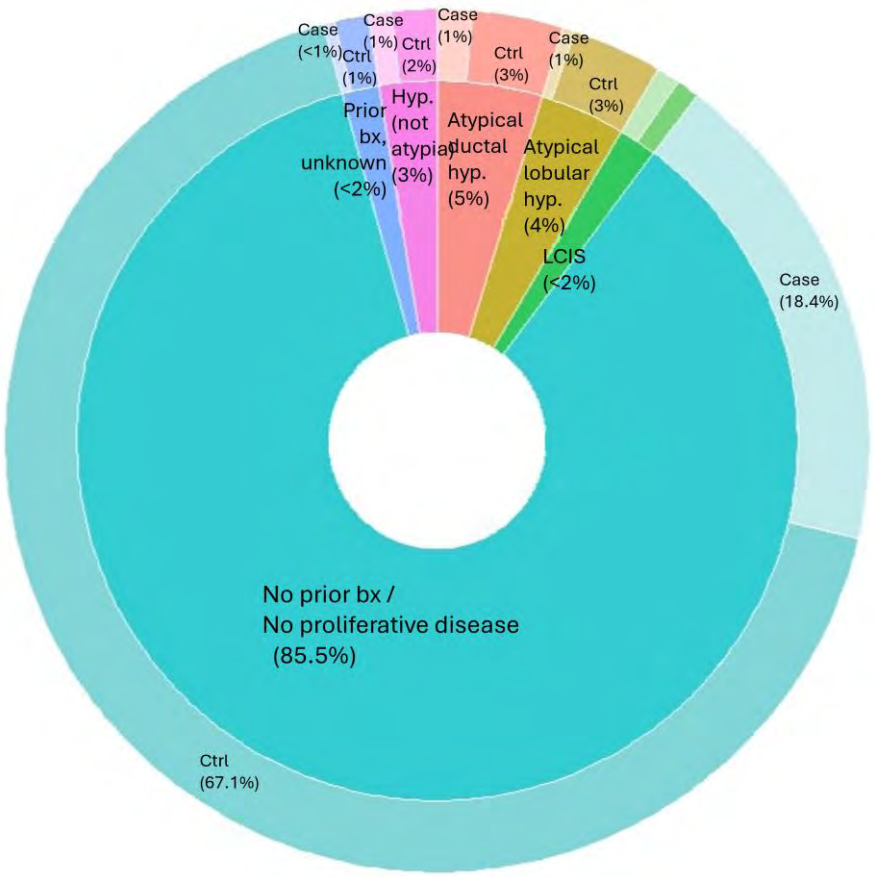

Figure 4.6
Pie chart for distribution of prior biopsy test results, along with breakdown for positives (Case) and negatives (Ctrl).

Family History of Breast Cancer

The largest group is for those who have at least one first degree relative and at least one second degree relative, making up 31% of the study cohort. Proportionally, the proportion with first degree relatives only (25%) is quite similar to second degree only (21%).

	Key	Total (n=234)	Cases (n=52)	Ctrl (n=182)
At least one first degree & at least one second degree	"1st Deg >= one, 2nd Deg >= one"	74 (32%)	26 (50%)	48 (27%)
First degree only	"1st Deg Only, one"	51 (22%)	5 (9%)	46 (25%)
Second degree only	"2nd Deg Only, >= one"	48 (21%)	13 (25%)	35 (19%)
Third degree only	"3rd Deg Only, >= one"	6 (3%)	-	6 (3%)
At least two first degree	"1st Deg, >= two"	10 (4%)	1 (2%)	9 (5%)
Has no family history	"No Fam Hx"	43 (18%)	6 (12%)	37 (20%)
Unknown	"Unknown"	2 (<1%)	1 (2%)	1 (<1%)

*All percentages calculated per column and rounded to the nearest whole number

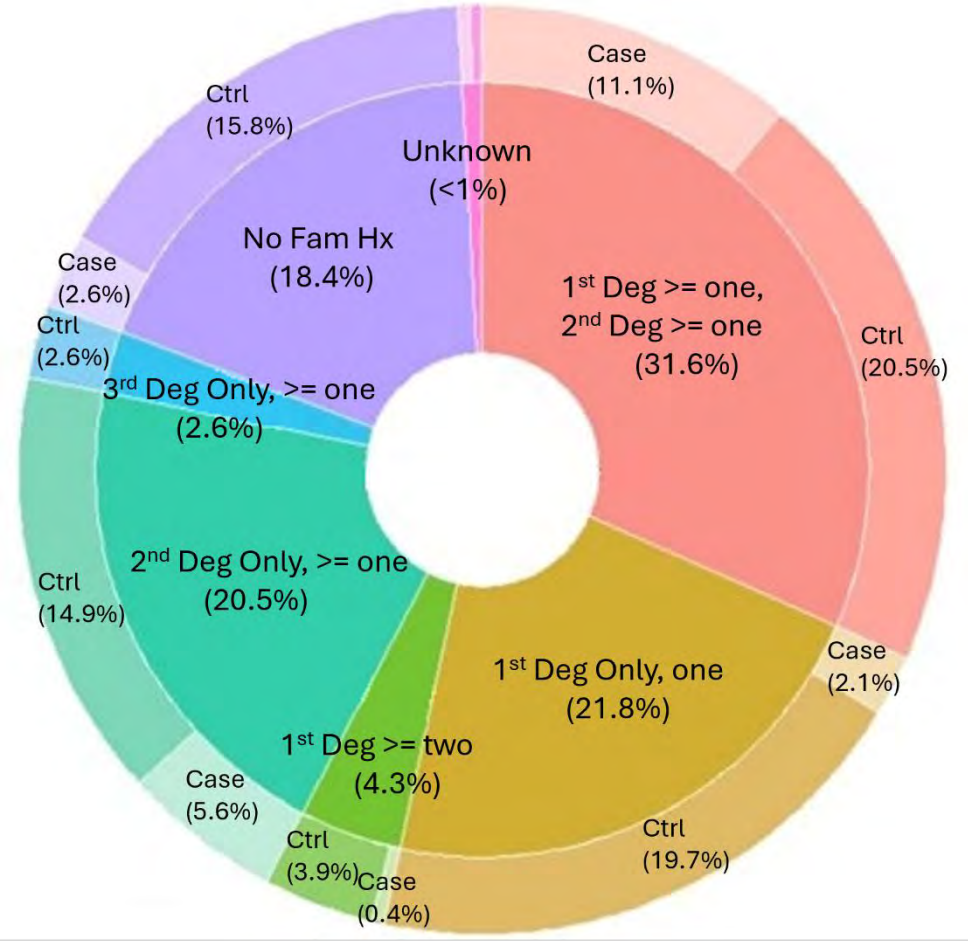

Figure 4.7
Pie chart for distribution of family history of breast cancer, along with breakdown for positives (Case) and negatives (Ctrl).

Table 4.2 summarizes the patient characteristics of the study cohort. Our cohort consists of 234 women (median age: 46, range: 22-77)), mostly premenopausal (59%), almost three-quarters being heterogeneously or extremely dense (72%) and more than half with BPE reported as minimal or mild (63%). BRCA test results were unknown for 73% of our cohort. Of the 52 women who developed cancer within five years (22%), 26 were invasive ductal carcinoma, 20 were ductal carcinoma in situ and 6 were invasive lobular carcinoma.

Table 4.2 Patient characteristics of the study cohort

Variable	**Categories**	**Study Cohort (N = 234)**
Age at MRI, years		46 (22-77)
Menopausal status		
	Pre-menopausal	138 (59%)
	Peri-menopausal	3 (1.3%)
	Post-menopausal	86 (36.7%)
	Missing	7 (3%)
Breast Density *		
	Fatty	7 (3%)
	Scattered	58 (25%)
	Heterogeneously dense	114 (49%)
	Extremely dense	55 (23%)
Qualitative BPE		
	Minimal	75 (32%)
	Mild	72 (31%)
	Moderate	53 (22%)
	Marked	34 (15%)
BRCA Test Results		
	BRCA 1	28 (12%)
	BRCA 2	17 (7%)
	Negative	18 (8%)
	Missing	171 (73%)
Tumor histology		
	Invasive ductal carcinoma (IDC)	26 (11%)
	Invasive lobular carcinoma (ILC)	6 (2%)
	Ductal carcinoma in situ (DCIS)	20 (9%)
	NA (Controls)	182 (78%)

All data reported in count (%) or median (interquartile range). * Density based on MR or mammography.

Most of the risk factors we collected had good completeness, with BRCA test results being the only exception, being unknown in 73% of the cohort. This is somewhat expected, due to the fact that BRCA testing is not a standard procedure performed on most women who receive cancer screening.

Collecting this information allows us to 1) calculate TC risk scores ourselves while being aware of the extent of data missingness and 2) perform subset analyses through selecting patients based on specific characteristics (see Chapter 6. Section 6.3). BRCA test results are expected to greatly influence TC risk score calculations for individuals who have them. The findings summarized in this chapter gives us an important perspective on the availability of BRCA test results when examining risk assessment tools that consider this data field.

In summary, we successfully built a high-quality database which we can leverage for the purposes of this dissertation. Clinical risk factor collection was performed to the best of our ability, considering the challenges in data collection. From our findings, we tabulated the different categories for every clinical risk factor collected and reported the extent of missingness. With the available clinical risk factors, we are well positioned to calculate TC risk scores to be used for multimodal prediction alongside quantitative imaging risk markers in Chapter 6. Before we perform the multimodal work, we need to establish the predictive performance of the quantitative imaging risk markers, which is the focus of Chapter 5.

Chapter 5

Quantitative MR imaging markers and cancer risk

In this chapter, we mainly explore prediction of cancer risk using quantitative imaging markers. In Section 5.2, we explore the predictive performance of quantitative imaging markers with cancer diagnosis within 5 years. This is enabled by having 5-year cancer outcomes, with case-cohort sampling to produce a dataset size feasible for extracting quantitative imaging markers and performing risk factor lookup, as described in Chapter 4. This same cohort will be the basis for much of the work in Chapter 6.

Since our work focuses on predicting future cancer, all the MRI images used are negative screening scans. We hypothesize that among all negative scans, those who develop cancer in the future may exhibit subtle differences in their imaging markers when compared to those who do not develop cancer. These differences may provide early indication of pre-cancerous changes in breast tissue, offering a direct, personalized reflection of the individual's biology and cancer risk.

5.1 BPE as a marker for cancer risk

Background parenchymal enhancement (BPE), which reflects the extent of fibroglandular tissue (FGT) enhancement on post-contrast dynamic contrast-enhanced (DCE) MR images, has shown promise as an imaging marker of breast cancer risk. Radiologists assess BPE qualitatively on DCE subtraction images (first post-contrast minus pre-contrast) and maximum intensity projections (MIPs) using the four categories of minimal, mild, moderate and marked, in accordance to BI-RADS guidelines (D'Orsi et al., 2013). Even though the exact mechanism by which BPE influences breast cancer development remains unclear, it is thought that BPE reflects areas of physiologically active tissue that responds to hormonal fluctuations. Estrogen is a steroid hormone with known histamine-like effects that can cause vasodilation and increased permeability of vessels (Zeppa, 1969), and the increased leakage of contrast material by the fibroglandular tissue is then observed as increased BPE (C. K. Kuhl et al., 1997). The tissue regions with elevated BPE indicate areas of physiologically active breast tissue, which are perhaps prone to increased inflammation and provide favorable

microenvironments for tumorigenesis (Bhatelia et al., 2014). While further studies are needed to explicitly interrogate the biological mechanism linking BPE and risk of breast cancer development, studies have established the association between qualitative BPE and breast cancer risk, independent of breast density (Arasu et al., 2019; Dontchos et al., 2015; Hu et al., 2021; King et al., 2011; Telegrafo et al., 2016).

BPE can also be quantified as a series of metrics that capture the extent of enhancement in terms of intensity and volume observed between pre- and post-contrast images of a patient's breast tissue. While there is yet to be a consensus on optimal metrics including specific parameters, post-contrast sequence selection, and enhancement threshold choice, quantitative BPE metrics have been found to be associated with breast cancer risk.

This chapter describes optimization of quantitative BPE metrics as markers of breast cancer risk. In Section 5.2, we derive quantitative BPE metrics and evaluate association with future breast cancer outcomes in a screening population of 234 women, which after case-cohort sampling described in Chapter 4, includes 52 cases of cancer diagnosis within 5 years since the MRI, and 182 controls who do not have cancer within that period.

5.2 Quantitative fibroglandular tissue enhancement MRI markers associated with breast cancer risk in a 2005-2021 cohort with cancer outcomes (n=234)

Extending upon our previous work in Chapter 3 on quantitative markers, we explored the efficacy of quantitative BPE markers on a case-cohort sample drawn from a larger institutional cohort of women with negative MRI screening scans. While mammographic screening is offered to women at average risk of breast cancer, MRI screening is reserved for women at increased risk as determined through presence of family history and/or personal genetic mutation. Therefore, our study focuses on a higher risk cohort rather than patients at average risk. For the study cohort of 234 women, we have 5-year follow-up information which enables us to assess the performance of quantitative imaging markers to predict cancer diagnosis within 5 years instead of using the surrogate of TC risk group in Chapter 3.

5.2.1. Methods for studying the association of Quantitative BPE markers and breast cancer risk (n=234)

Patient Population and Clinical Factors

This study involved patients who underwent breast MRI screening due to being considered at elevated risk of breast cancer due to presence of a BRCA mutation, lifetime risk of 20% or greater according to Tyrer-Cuzick risk modeling, or family history of breast cancer. As outlined in Chapter 4, we began with a larger pool of patients who underwent screening between 2005 and 2021, then excluded individuals with a personal history of

prior breast cancer and those with cancer diagnosis within 1 year after the screening MRI. An enriched case-cohort was created to reduce processing burden while maintaining statistical power, which included a total of 234 women: 52 cases diagnosed with cancer within five years after MRI and 182 controls confirmed to be negative for cancer based on biopsy and 1-year follow-up.

For the 234 patients in our study cohort, we collected the core clinical risk factors outlined in Chapter 4, wherever available, from electronic medical records: age at screening MRI, age at menarche, menopausal status at MRI, age at menopause if postmenopausal, breast density, BRCA genetic test result, personal history of breast biopsy, and family history of breast cancer. These clinical risk factors were used to calculate 5-year Tyrer-Cuzick lifetime risk scores using the IBIS Breast Cancer Risk Evaluation Tool (CharterHouse Square, London, United Kingdom). Missing risk factors were treated as “N/A” in the calculator tool.

Image Acquisition and Markers

Since our study cohort scans spanned 2005 to 2021, patients were scanned on one of the scanners listed below in the Table:

Table 5.1. Breakdown of acquisition scanner systems and cases vs controls for the n=234 study cohort

Scanner System	Field Strength (T)	Cases	Controls
Philips Achieva	3	29	125
Philips Ingenia	1.5	1	0
GE SIGNA EXCITE	1.5	14	31
GE SIGNA HDx	1.5	5	19
GE SIGNA HDxt	1.5	3	7

The scanner systems involved and their acquisition time periods are as follows: Philips Achieva 3T (February 2010 – March 2021), Philips Ingenia 1.5T (October 2019), GE SIGNA EXCITE (January 2005 – October 2007), GE SIGNA HDx (October 2007 – January 2009), GE SIGNA HDxt (March 2009 – January 2010). Dynamic contrast-enhanced (DCE) MRI involved one pre-contrast and multiple post-contrast T1-weighted scans, with the first post-contrast scan centered at 110 s after contrast injection (0.1 mmol/kg body weight gadolinium-based contrast agent).

Qualitative BPE assessments were obtained from imaging reports, where BPE was assessed by radiologists using DCE subtraction images (first post-contrast minus pre-contrast) and maximum intensity projections (MIPs) at the time of clinical interpretation, and categorized as minimal, mild, moderate, or marked in accordance with BI-RADS guidelines (D’Orsi et al., 2013).

Quantitative BPE derivation methods are the same as in Chapter 3, except that Section 5.2 represents a subset of four quantitative metrics, as outlined in Table 5.2.

Briefly, custom software developed in MATLAB (R2021b 9.11.0.2358333 Update 7; Mathworks, Natick, MA, USA) was used to calculate quantitative markers. Pre-contrast DCE images underwent preprocessing to correct for patient motion using 2D affine registration and then resampled to isotropic resolution (1×1×1 mm). Image segmentation was performed using an open-source deep learning-based algorithm. Volumetric masks were generated for FGT and used to derive volumetric metrics for the analysis (Table 5.2).

Table 5.2. Quantitative BPE and volume measures calculated for the n=234 study cohort.

	Derivation Formula
Volume measures	
FGT volume (cm^3)	# of voxels in FGT mask × volume per voxel
BPE measures	
Median PE (%)	median PE of voxels > 10% within FGT mask
BPE volume (cm^3)	# of voxels with PE > 0% within FGT mask × volume per voxel
BPE:FGT ratio (%)	(BPE volume/FGT volume) × 100%
Integrated intensity (cm^3)	BPE volume × mean PE

Statistical Analyses

The primary analysis involved logistic regression models to assess the associations of individual imaging markers with binary cancer outcome (labeled as “0” or “1”) within 5 years, on a 10-fold cross-validation loop. In this initial analysis, a simplified approach was taken to analyze the data in the manner of a case-control design, ignoring the time-to-event information. To fully utilize the time-to-event information, a more complex time-varying weighted Cox regression is required, beyond the scope of this dissertation. Performance of predicting cancer outcome status using individual imaging markers was assessed using area under the receiver operating characteristic curve (AUC). Histograms were used to visually assess the distributions of markers across patients. Markers demonstrating right-skewness were log-transformed to reduce asymmetry before including them in the models. This part of the thesis document will discuss the performance of the quantitative imaging markers and compare them to qualitative BPE, while modeling involving TC risk score will be included later in Chapter 6.

Optimization of Quantitative BPE Marker Calculation

Since our study cohort’s images were acquired on multiple scanner systems and field strengths, there may be concerns over 3T images having more pronounced signal intensity and ranges than 1.5T images, ultimately affecting the numerical comparisons in instances when cases and controls are of different field strengths. Therefore, instead of selecting a single enhancement threshold for the quantitative BPE metrics, we separated patients scanned on 1.5T vs 3T and derived optimized thresholds for each field strength during every fold of the model training phase. For patients of the same field strength, we optimized selection by running all nine thresholds of 10% to 90% for each quantitative

BPE measure during model training and selecting one with the highest AUC as the optimized threshold to use in the subsequent model testing. In the resulting data frame, patients with the same field strength will have quantitative markers set at the same optimized threshold in that fold. This is illustrated in a table below showing optimized threshold results selected during model training, where rows are color-coded by the specific quantitative marker. Note that in this particular fold, for a given marker, for instance median PE, the optimized threshold level is found to be 30% at 1.5T, but 60% at 3.0T. This optimization strategy allows us to select enhancement threshold levels that are best at distinguishing between cases and controls, while reducing the influence of field strength since the thresholds are selected separately in the process.

Table 5.3. A sample table of optimized thresholds for quantitative BPE metrics selected from a single fold

Field Strength (T)	Quantitative Marker	Optimized Threshold level
1.5	Median PE (%)	30%
	BPE volume (cm^3)	60%
	BPE:FGT ratio (%)	10%
	Integrated intensity (cm^3)	60%
3.0	Median PE (%)	60%
	BPE volume (cm^3)	50%
	BPE:FGT ratio (%)	70%
	Integrated intensity (cm^3)	10%

5.2.2. Results for studying the association of optimized Quantitative BPE markers and breast cancer risk (n=234)

Patient Characteristics and Follow-Up

Our study cohort involved 234 women (median age: 46 years, range: 22–77 years), of which 52 (22%) developed cancer 1 to 5 years after index MRI with median time to diagnosis of 2.7 (range: 1.0 to 4.8) years. The majority of women were pre- or peri-menopausal (60.3%), did not have a known BRCA mutation (73%), and exhibited minimal or mild (63%) qualitative BPE (Table 4.9). The proportion of women with moderate or marked qualitative BPE was greater in the cases group (25/52; 48%) compared to the controls (62/182; 34%), which aligns with associations of higher qualitative BPE with risk.

Table 5.4. Patient characteristics for n=234 study cohort.

Variable		Overall Cohort (N = 234)
Age at MRI, years		46 (39, 54)
Menopausal status		
	Pre-menopausal	138 (59%)
	Peri-menopausal	3 (1.3%)
	Post-menopausal	86 (36.7%)
	Unknown	7 (3%)
BRCA mutation		
	BRCA 1	28 (12%)
	BRCA 2	17 (7%)
	Negative	18 (8%)
	Unknown	171 (73%)
Qualitative BPE		
	Minimal	75 (32%)
	Mild	72 (31%)
	Moderate	53 (22%)
	Marked	34 (15%)

All data reported in count (%) or median (interquartile range).

Association between Quantitative BPE metrics and 5-Year Cancer Outcome

The AUCs for univariate modeling of the four quantitative BPE metrics and qualitative BPE are as follows: BPE Volume, 0.595; BPE:FGT Vol, 0.622; Integrated Intensity, 0.461; Median PE, 0.543; qualitative BPE 0.550. Three out of four quantitative BPE metrics performed comparably close or better than qualitative BPE. The ROC curves for these are displayed in Figure 5.1.

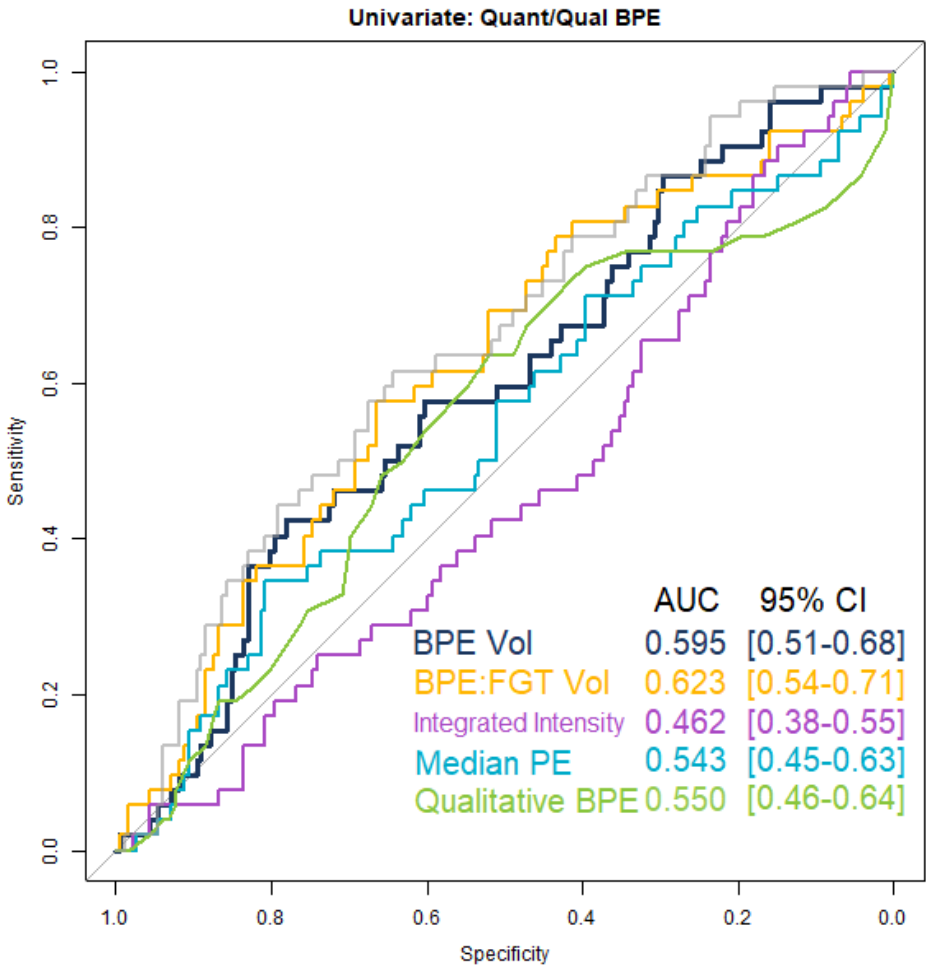

Figure 5.1. Diagnostic performance for predicting cancer diagnosis ≤ 5 years after MRI using quantitative BPE markers and qualitative BPE.

5.2.3. Discussion for studying the association of Quantitative BPE markers and breast cancer risk (n=234)

Results from our univariate analysis showed that quantitative BPE has a modest improvement of classification performance over qualitative BPE. Two markers, BPE Vol and BPE:FGT Vol (AUCs 0.595 and 0.623, respectively) are numerically higher than qualitative BPE (AUC: 0.55). This generally aligns with findings from studies which found that quantitative BPE outperform radiologist-assessed qualitative BPE in associating with breast cancer risk (Lam et al., 2019; Niell et al., 2021). A clear advantage that quantitative measures have over qualitative ones is that they are less susceptible to inter-rater variability, and due to being derived from an image processing pipeline, represent more objective and reproducible markers of cancer risk. We believe that quantitative metrics derived from voxel-wise changes between pre- and post-contrast captures a far more granular measure of fibroglandular tissue state than qualitative assessments of overall

enhancement across the entire bilateral breasts. This suggests an opportunity to improve conventional risk models by incorporating quantitative BPE markers which may capture information about the patient's tissue biology more directly. Furthermore, since conventional risk models do not incorporate imaging-derived information beyond overall mammographic density, quantitative markers may provide complementary information to the clinical variables included in such risk models.

There are several limitations to this section of our results. Since our study cohort involved women who underwent screening MRI due to being at high risk of breast cancer, our findings may limit the generalizability of our findings to a broader population of average risk women. Qualitative BPE was assessed manually by multiple radiologists and would be more subject to inter-rater viability than if they were assessed by a single reader. Additionally, we did not control for menstrual cycle or hormonal therapy which may impact BPE levels, since that information was not available to us.

5.2.4. Conclusion for studying the association of Quantitative BPE markers and breast cancer risk (n=234)

We investigated the ability of quantitative BPE markers to predict cancer within 5 years in women who are at high risk of breast cancer. Our findings suggested that some quantitative BPE markers were moderately more predictive than qualitative BPE, while providing objective, automated assessment. This presents the potential to improve conventional clinical risk models by incorporating quantitative BPE markers that may provide incremental value. The next steps would involve exploring models that incorporate both quantitative BPE and conventional risk model scores, which will be the main topic of discussion in the next chapter.

Chapter 6
Multimodal combination of imaging and clinical risk factors (n=234)

In Chapter 5 we presented results of cancer risk prediction using imaging markers alone. In this chapter, we explore improving cancer risk prediction by combining quantitative imaging markers with conventional TC risk scores. An important motivation for our work in cancer risk prediction is the incorporation of both imaging and clinical variables during modeling. Since TC risk scores are calculated using clinical factors, they may potentially provide different but complementary information about an individual's cancer risk. My hypothesis is that combining these with qualitative and quantitative BPE markers may result in a more holistic representation of risk that performs better than using just the clinically-derived TC risk scores alone.

Section 5.2 involves univariate modeling of quantitative and qualitative BPE, while Section 6.2 presents a bivariate, multimodal approach of adding TC risk scores as a covariate to each quantitative and qualitive BPE model.

6.1 Methods for multimodal combination of imaging and clinical risk factors (n=234)

Patient Population & Image Acquisition and Markers

As mentioned previously, the cohort used here is the same as presented in Chapter 4 and used in Chapter 5. There is a total of 234 women, of which 52 were cancer cases that developed within five years from the time of MRI scanning. The proportion of cancer development is high due to the case-cohort sampling approach described in Chapter 4. A summary of the patient characteristics can be found in Table 5.4.

Based on the work I carried out and reported in Chapter 4, we used the stored clinical risk factors to calculate TC risk scores using the IBIS Breast Cancer Risk Evaluation Tool (CharterHouse Square, London, United Kingdom). Missing risk factors were treated as "N/A" in the calculator tool. Chapter 4 offers a more detailed look at data

missingness for the various clinical factors we collected. We used insights from some of the clinical factors collected to perform analyses of subsets from the cohort of 234 individuals.

Since the patient population use here is the same as in Sections 5.2, the image acquisition details and quantitative marker derivation are the same as described before.

Statistical Analysis

The statistical analysis here has a similar approach as in Chapter 5, Section 5.2, where logistic regression models are used to assess the associations of individual imaging markers with cancer outcome (labeled as “0” or “1”) within 5 years, on a 10-fold cross-validation loop. Prediction was performed using univariate and bivariate approaches. Univariate models involve qualitative BPE, breast density, TC risk score, and four quantitative BPE measures. The univariate prediction results for BPE metrics are reported in Chapter 5, Section 5.2, while the multimodal prediction results are reported in the current chapter.

Next, we present bivariate, multimodal models incorporating imaging information (qualitative BPE, quantitative BPE, density) and clinical information (TC risk score).

Same as before, quantitative BPE metrics were log2 transformed prior to statistical analyses. We assessed the association with 5-year breast cancer risk using area under the receiver operator characteristic curve (AUC), differentiating women developing cancer within 5 years from those without cancer. Logistic regression models leveraged TCv8 risk scores, qualitative BPE and quantitative BPE metrics as predictor variables in different variants. All numerical data underwent z-score normalization, and all classifier variants underwent 10-fold cross validation for hyperparameter tuning. Performance for predicting 5-year breast cancer was assessed using the AUC, estimated using 10-fold cross-validation.

Aside from AUC, I also examined model performance through the lens of sensitivity and specificity. To evaluate the performance of prediction models for clinical implementation, sensitivity and specificity are more relevant metrics. Sensitivity and specificity allow for the consideration of false positives and false negatives as tradeoffs to balance when evaluating the performance of a model during as part of clinical decision making. As an exploratory approach, I investigate sensitivity when specificity is defined at 90%, and specificity when sensitivity is defined at 90%. This allows me to compare performance across the various multimodal combinations at the referenced levels of sensitivity and specificity. In Section 6.2, for simplicity of result presentation, I focus only on the full cohort (n=234) and compare sensitivity and specificity between the TC Only model and the highest-performing multimodal model by AUC.

6.2 Results for Multimodal combination of imaging and clinical risk factors (n=234)

The AUCs for bivariate modeling of the four quantitative BPE metrics and qualitative BPE are as follows: BPE Volume, 0.677; BPE:FGT ratio (Vol), 0.673; Integrated Intensity, 0.637; Median PE, 0.642; qualitative BPE, 0.674.

Models for two out of four quantitative BPE metrics performed comparably or better than qualitative BPE. The ROC curves for these are displayed in Figure 6.1. In contrast, univariate prediction in this cohort using TC risk score has an AUC of 0.648.

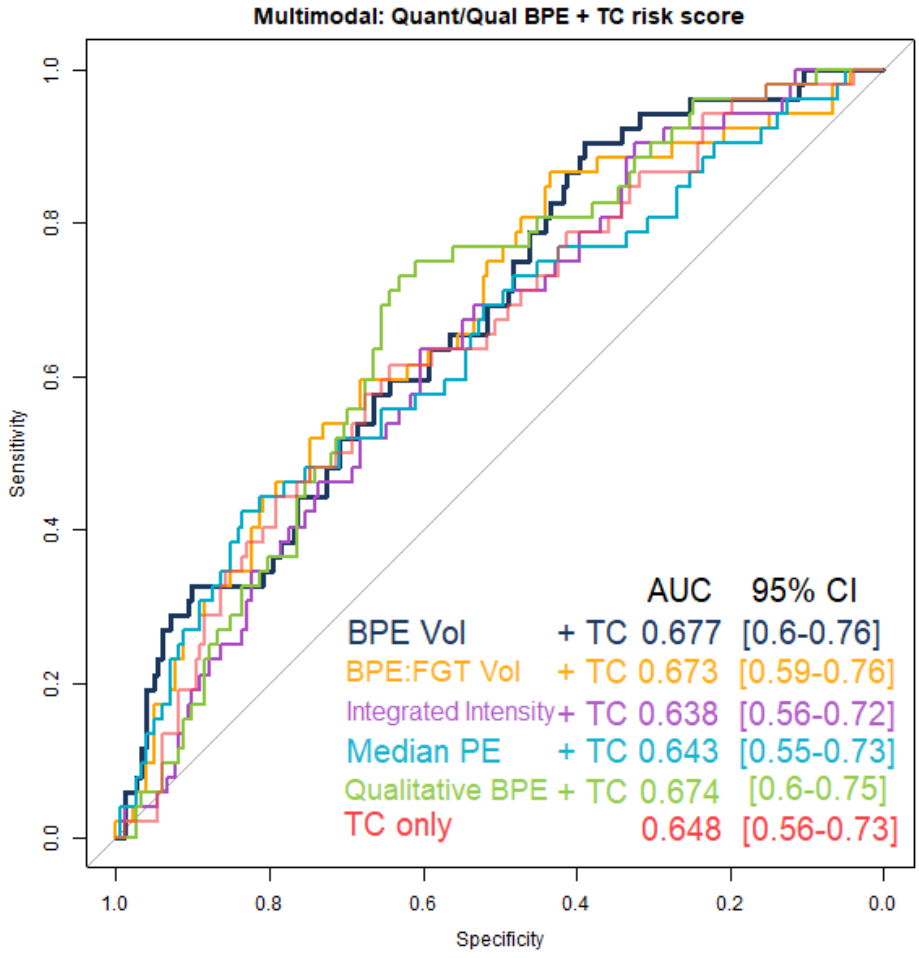

Figure 6.1. Performance of multimodal combination of clinical TC risk scores and imaging markers to predict cancer diagnosis within 5 years.

Since BPE Volume + TC achieved the highest AUC (0.677) out of all multimodal combinations, here I compared it against TC Only in terms of sensitivity and specificity.

In Figure 6.2, dashed lines in shades of blue pertain to BPE Volume + TC, and dashed lines in shades of red pertain to TC Only. When sensitivity is set at 90%, specificities are 39% and 24.2% for BPE Volume + TC and TC Only, respectively. When specificity is set at 90%, sensitivities are 32.7% and 19.2% for BPE Volume + TC and TC Only, respectively. The multimodal combination of BPE Volume + TC outperforms TC Only in both sensitivity and specificity.

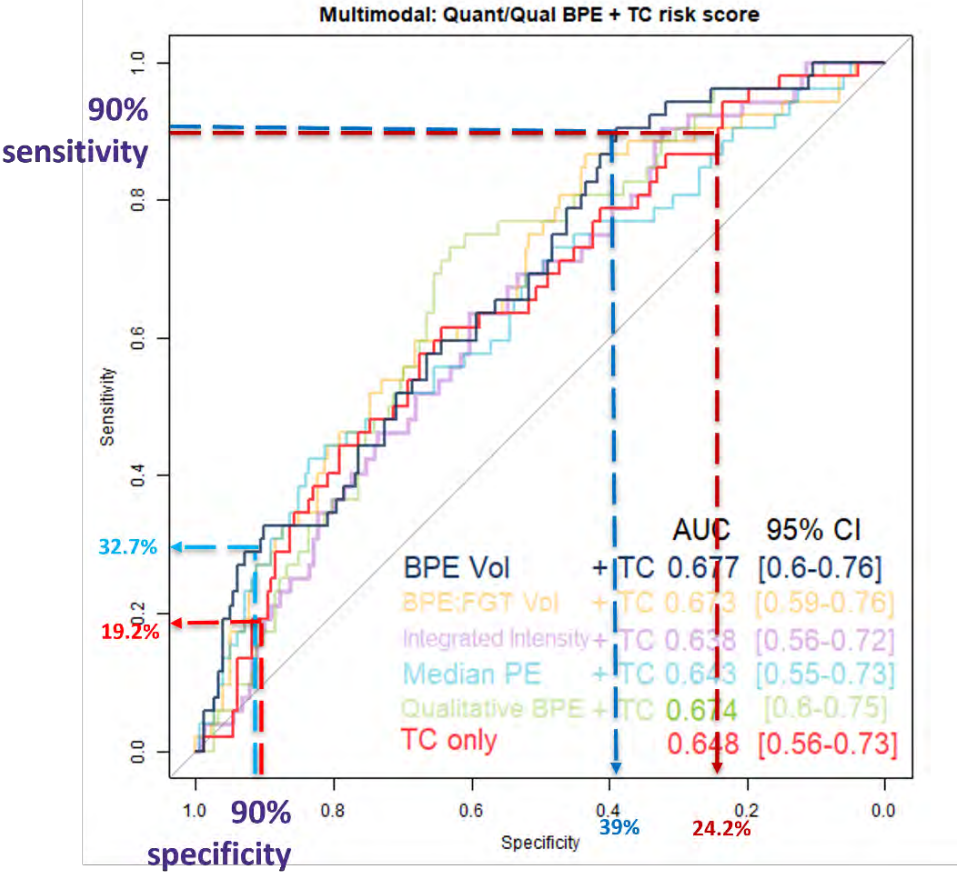

Figure 6.2. Comparing levels of sensitivity and specificity between BPE Volume + TC and TC Only, for predicting cancer diagnosis within 5 years.

6.3 Subset analysis

As part of the multimodal analyses, we performed several subgroup analyses. The subgroups were chosen based on known clinical factors expected to influence imaging-based prediction efficacy. A breakdown of cases and controls for each subset is summarized in Table 6.1, and comparison of multimodal modeling predictive performance across the subsets is reported in Table 6.2.

Table 6.1 Breakdown of cases and controls in the whole cohort and subsets

Cohorts	N	Cases	Controls
Full Cohort (6.2)	234	52	182
No Known BRCA mutation (6.3.1)	182	31	151
Invasive Cancers Only (6.3.2)	212	32	180
Pre-menopausal (6.3.3)	138	35	103

Table 6.2 Comparison of multimodal prediction performance across cohort subsets: AUC, sensitivities, specificities

		Full Cohort N=234	No known BRCA mutation N=182	Invasive Cancer Only N=212	Pre-Menopausal N=138
TC Only	AUC (95% CI)	0.648 (0.564-0.731)	0.611 (0.506-0.715)	0.609 (0.501-0.716)	0.693 (0.591-0.793)
	Sensitivity at 90% specificity	19.2%	19.4%	25%	37.1%
	Specificity at 90% sensitivity	24.2%	25.8%	17.8%	25.2%
Qualitative BPE + TC	AUC (95% CI)	0.674 (0.596-0.751)	0.696 (0.603-0.789)	0.641 (0.54-0.742)	0.718 (0.626-0.808)
	Sensitivity at 90% specificity	17.3%	22.6%	18.8%	31.4%
	Specificity at 90% sensitivity	30.2%	39.1%	25.6%	29.1%
Median PE + TC	AUC (95% CI)	0.643 (0.554-0.731)	0.664 (0.561-0.767)	0.656 (0.558-0.753)	0.663 (0.559-0.765)
	Sensitivity at 90% specificity	26.9%	32.3%	15.6%	25.7%
	Specificity at 90% sensitivity	22.0%	35.8%	28.9%	26.2%
BPE Volume + TC	AUC (95% CI)	0.677 (0.598-0.756)	0.664 (0.57-0.757)	0.613 (0.507-0.718)	0.672 (0.571-0.771)
	Sensitivity at 90% specificity	32.7%	22.6%	18.8%	28.6%
	Specificity at 90% sensitivity	39.0%	34.4%	13.3%	32.0%
BPE:FGT ratio + TC	AUC (95% CI)	0.673 (0.59-0.756)	0.672 (0.573-0.769)	0.575 (0.458-0.691)	0.677 (0.573-0.78)
	Sensitivity at 90% specificity	26.9%	25.8%	18.8%	34.3%
	Specificity at 90% sensitivity	27.5%	25.2%	11.1%	23.3%
Integrated Intensity + TC	AUC (95% CI)	0.638 (0.556-0.718)	0.656 (0.561-0.75)	0.625 (0.519-0.731)	0.679 (0.578-0.778)
	Sensitivity at 90% specificity	19.2%	12.9%	15.6%	20%
	Specificity at 90% sensitivity	32.4%	31.1%	14.4%	32.0%

6.3.1 No Known BRCA mutations (n=182)

Here we excluded individuals with known BRCA mutation. Since a positive BRCA test result outweighs other risk factors from modeling, investigating this subset enables us to focus on individuals who will benefit the most from risk prediction; women who have negative or unknown BRCA test results. Additionally, it was previously shown that BRCA mutation carriers, who are at extremely high risk, tend to paradoxically exhibit lower qualitative BPE than non-carriers (Grubstein et al., 2018). Excluding BRCA-positive individuals would theoretically reduce instances where the BRCA-driven risk contradicts the BPE-level associated risk and improve risk prediction in this subset.

For individuals with no known BRCA mutations, the AUCs for bivariate modeling of the four quantitative BPE metrics and qualitative BPE are as follows: BPE Volume, 0.664; BPE:FGT ratio (Vol), 0.671; Integrated Intensity, 0.655; Median PE, 0.664; qualitative BPE, 0.696. Here, qualitative BPE achieved a higher AUC than those of quantitative BPE metrics. In comparison, univariate prediction in this cohort using TC risk score has an AUC of 0.61.

I compared sensitivities and specificities between TC Only and the combination with the highest AUC (0.696), which is qualitative BPE + TC. When sensitivity is set at 90%, specificities are 39.1% and 25.8% for qualitative BPE + TC and TC Only, respectively. When specificity is set at 90%, sensitivities are 22.6% and 19.4% for qualitative BPE + TC and TC Only, respectively. The multimodal combination of qualitative BPE + TC outperforms TC Only in both sensitivity and specificity.

6.3.2 Invasive Breast Cancer only (No DCIS) (n=212)

Here we excluded individuals whose cancer diagnosis was ductal carcinoma in situ (DCIS) instead of invasive carcinoma. Qualitative BPE associates more strongly with invasive cancer than with DCIS (Arasu et al., 2019), possibly due to the fact that DCIS tends to more likely manifest radiologically as a non-mass (Rosen et al., 2007). This leads us to hypothesize that predicting future DCIS cancer from negative screening scans would be especially difficult and therefore focus on predicting future invasive cancer in this sub-analysis.

By restricting positive cases to those developing invasive cancers, the AUCs for bivariate modeling of the four quantitative BPE metrics and qualitative BPE are as follows: BPE Volume, 0.612; BPE:FGT ratio (Vol), 0.575; Integrated Intensity, 0.625; Median PE, 0.656; qualitative BPE, 0.641. The model for median PE achieved a numerically higher AUC than qualitative BPE. In comparison, univariate prediction in this cohort using TC risk score has an AUC of 0.608.

I compared sensitivities and specificities between TC Only and the combination with the highest AUC (0.656), which is median PE + TC. When sensitivity is set at 90%, specificities are 28.9% and 17.8% for median PE + TC and TC Only, respectively. When

specificity is set at 90%, sensitivities are 15.6% and 25% for median PE + TC and TC Only, respectively. The multimodal combination of median PE + TC outperforms TC Only in specificity.

6.3.3 Pre-menopausal only (n=138)

Here we excluded individuals known to be peri- or post-menopausal due to the fact that qualitative BPE was found to be more strongly associated with cancer risk for premenopausal women than for postmenopausal women (Watt et al., 2020).

When looking only at pre-menopausal individuals, the AUCs for bivariate modeling of the four quantitative BPE metrics and qualitative BPE are as follows: BPE Volume, 0.671; BPE:FGT ratio (Vol), 0.677; Integrated Intensity, 0.678; Median PE, 0.662; qualitative BPE, 0.717. Here, qualitative BPE achieved a higher AUC than those of quantitative BPE metrics. In comparison, univariate prediction in this cohort using TC risk score has an AUC of 0.692.

I compared sensitivities and specificities between TC Only and the combination with the highest AUC (0.717), which is qualitative BPE + TC. When sensitivity is set at 90%, specificities are 29.1% and 25.2% for qualitative BPE + TC and TC Only, respectively. When specificity is set at 90%, sensitivities are 31.4% and 37.1% for qualitative BPE + TC and TC Only, respectively. The multimodal combination of qualitative BPE + TC outperforms TC Only in specificity.

6.4 Discussion for multimodal combination of imaging and clinical risk factors

In examining multimodal prediction of cancer across various cohort subsets, we found that quantitative BPE metrics perform as well or better than qualitative BPE in our full cohort (n = 234) and invasive-cancers subset (n = 212). This finding helps strengthen the point that since quantitative BPE metrics are more objectively derived, they are therefore more robust risk markers when compared to qualitative BPE which is more susceptible to inter-reader variability.

The second important finding from our results is the incremental value that imaging markers hold relative to using clinical TC risk scores alone. Table 6.2 summarizes this, showing that for every cohort subset there are models combining imaging and TC risk score which outscore the univariate model for TC risk score only.

In the full cohort of 234 individuals, we see that two quantitative BPE models (BPE Volume + TC and BPE:FGT ratio (Vol) + TC) perform comparably or better (AUCs: 0.677 and 0.673, respectively) than qualitative BPE (AUC: 0.674), and all these numerically outperform TC Only (AUC: 0.648) despite similarities in confidence interval. This pattern is also similarly observed in the cohort subset of 212 women which includes only invasive cancers as positives. A quantitative BPE model, median PE + TC (AUC: 0.656), performs slightly better than qualitative BPE + TC (AUC: 0.641), and both outperform TC Only (AUC:0.608). As reported by others (Lam et al., 2019; Niell et al., 2021) and our own work (Surento et al., 2025), quantitative BPE metrics at optimal thresholds have the capacity

to outperform qualitative BPE in predicting cancer risk. While the magnitude of outperforming is less pronounced in our findings, they nevertheless provide new evidence that this may still be true even with the addition of a clinical factor in the form of TC risk scores. Even when quantitative BPE metrics perform comparably similar to qualitative BPE, they are still favored due to being more reproducible and objective markers than manually-assessed qualitative BPE categories. Performance by sensitivity and specificity (see Table 6.2 & Figure 6.2) showed that the best-performing multimodal combination by AUC, BPE Volume + TC, outperformed TC Only.

In the cohort subset of 182 individuals where BRCA positive women were excluded, we observe that models of quantitative BPE + TC model were outperforming the univariate TC Only model but did not outperform qualitative BPE + TC. Individuals with BRCA mutation have significantly increased cancer risk but also paradoxically can exhibit lower qualitative BPE, so it is expected that removing them would produce an improvement in qualitative BPE prediction performance. Models of quantitative BPE metrics showed a stabilizing effect, where the range of AUC (lowest to highest: 0.655 to 0.671) is a narrow difference of 0.016. This finding suggests that among women who do not have BRCA mutation but develop cancer in the future, quantitative BPE may not have captured differences as holistically as qualitative BPE which may include aspects of FGT beyond BPE, as well as BPE distribution patterns not captured by the quantitative metrics. This subset of 182 women is also the only one in our study where all models of imaging (qualitative or quantitative BPE + TC) outperform (AUCs: 0.655-0.696) the univariate TC Only model (AUC: 0.61). This illustrates that the presence of BRCA mutation greatly influences a person's risk and the prediction performance of their TC risk score. As discussed in Chapter 4, since a BRCA mutation test result commonly override risk scores from modeling, this analysis enables us to focus on other individuals who would benefit the most from risk modeling; women who have negative or unknown BRCA test results. The results generally show that adding imaging markers outperform using TC Only, thus further strengthening the value of using imaging markers to supplement TC risk score for risk prediction. Performance by sensitivity and specificity (see Table 6.2) showed that the best-performing multimodal combination by AUC for this subset, qualitative BPE + TC, outperformed TC Only.

In the subset of 212 women with only invasive cancers as positives, the numerical difference in AUC between the highest from median PE + TC (AUC: 0.656) and qualitative BPE +TC (AUC:0.641) is slightly more pronounced than in the full cohort where the difference between the best-performing BPE Volume + TC (AUC: 0.677) and qualitative BPE +TC (AUC: 0.674) is marginal. This is perhaps resulting from the fact that the cohort with 212 individuals only have invasive cancers as positives where DCIS cases were excluded; median PE may potentially associate with non-DCIS future cancers better by virtue of their enhancement characteristics. While our finding in this aspect may hint at this potential difference, it would need to be validated by future studies which focus on investigating the association between specific quantitative BPE metrics and cancer subtypes with specific enhancement characteristics. Performance by sensitivity and specificity (see Table 6.2) showed that the best-performing multimodal combination by AUC for this subset, median PE + TC, outperformed TC Only in specificity but not

sensitivity. This suggests that setting a high specificity of 90% results in median PE + TC to have a lower sensitivity than TC Only, which means more false negatives and not identifying enough future cases.

In the premenopausal subset of 138 women, qualitative BPE + TC achieved the highest AUC of 0.717, followed by the univariate TC Only at 0.692. Models of quantitative BPE + TC achieved AUCs of 0.662 to 0.678, while not numerically higher than qualitative BPE + TC or TC Only. The finding here still suggests that adding an imaging marker (qualitative BPE) provides incremental value to improve prediction performance relative to TC Only. Performance by sensitivity and specificity (see Table 6.2) showed that the best-performing multimodal combination by AUC for this subset, qualitative PE + TC, outperformed TC Only in specificity but not sensitivity. This suggests that setting a high specificity of 90% results in qualitative PE + TC to have a lower sensitivity than TC Only, which means more false negatives and not identifying enough future cases.

Overall, our results suggest that multimodal models generally improve upon univariate risk prediction. In the full cohort where there is more variability of patient characteristics and cancer subtypes than in cohort subsets, prediction may have benefited from synthesizing information from both imaging and clinical factors that explain different aspects of an individual's biology. Cohort subset analyses also demonstrated that cohort selection may influence the performance of TC risk score, and different imaging markers may excel in different scenarios. The magnitude of marker performance in subset analyses were also possibly constrained by the sample size of the full cohort from which they are drawn, assuming that larger cohorts contain more diversity of imaging and clinical characteristics, potentially conferring more robustness in findings. Information from sensitivities and specificities could be relevant for developing clinical applications of such prediction models, especially when considering the tradeoffs involving rates of false positives and negatives. Using the example of the full cohort of 234 women (see Table 6.2 and Figure 6.2), in the combination with the highest AUC (0.677), BPE Volume + TC, if a 90% sensitivity is prioritized, there may be fewer false negatives, at the cost of possibly including more false positives since the specificity is only at 39%. If 90% specificity is prioritized, there may be fewer false positives, at the cost of missing positives that get mis-labeled as negatives since the sensitivity is only 32.7%. Stakeholders such as clinicians, patients, and public health surveillance experts would be relevant to involve in deciding an optimized approach if such a model were to be incorporated as a screening tool.

6.5 Cohort subset analyses for predicting cancer diagnosis within 4 and 3 years

As part of exploratory analyses, I considered the risk predictions for cancer if the window for diagnosis were something other than 5 years. It is worth mentioning that when I also investigated prediction for 4 and 3 years, the cohort sizes were different from predicting 5-year cancer risk. For instance, someone who had cancer diagnoses at a later time point such as 4.5 years would not be considered as positive when predicting 3-year cancer risk because the cancer diagnosis is outside of the 3-year window. There was the option to either classify that individual as a negative within the 3-year window, or remove the individual entirely from the cohort. For the purposes of the exploratory analysis, I chose the latter to avoid any likelihood of including potential early tissue changes to be among the negatives. While the former option could be explored, they were not carried out and could be attempted for future work. Table 6.3 summarizes cohort size differences across subsets used for predicting each year of cancer risk. Figures 6.3 to 6.6 illustrate the AUC scores of each multimodal combination model across 5-, 4- and 3-year risk prediction.

Table 6.3 Cohort differences across prediction years of cancer risk

		Full Cohort
5-year cancer risk	Cases	52 (22.2%)
	Controls	182 (77.8%)
	Total	234 (100%)
4-year cancer risk	Cases	43 (19.1%)
	Controls	182 (80.9%)
	Total	225 (100%)
3-year cancer risk	Cases	33 (15.3%)
	Controls	182 (84.7)
	Total	215 (100%)

Multimodal combinations involving quantitative BPE metrics generally trend towards better AUC performance across the years of cancer risk. AUC values at 5-year cancer risk are higher than those at 4-year cancer risk, which are in turn higher than those at 3-year cancer risk. Qualitative BPE + TC deviates from that trend slightly, with 4-year cancer risk prediction performance being higher than at 5-year cancer risk. This was not an expected outcome, warranting further investigation to better understand the underlying reason. One hypothesis is that it may have to do with the differences between qualitative and quantitative BPE: qualitative BPE is a holistic assessment of enhancement that differs from the quantitative BPE metrics which are intended to capture individual aspects of enhancement. In order to better validate that line of thinking, it would require having sequential screening MRI scans for several years leading up to the cancer diagnosis for positives as well as for controls. Investigating qualitative and quantitative BPE for all those time points may potentially provide a clear progression of how these enhancement measures change during every year leading up to the cancer diagnosis as well as for those who do not develop cancer. Additionally, as shown in Table 6.3, the cohorts differ

at 5-year, 4-year, and 3-year cancer risk prediction, including number of positive cases, which may also influence model prediction performance.

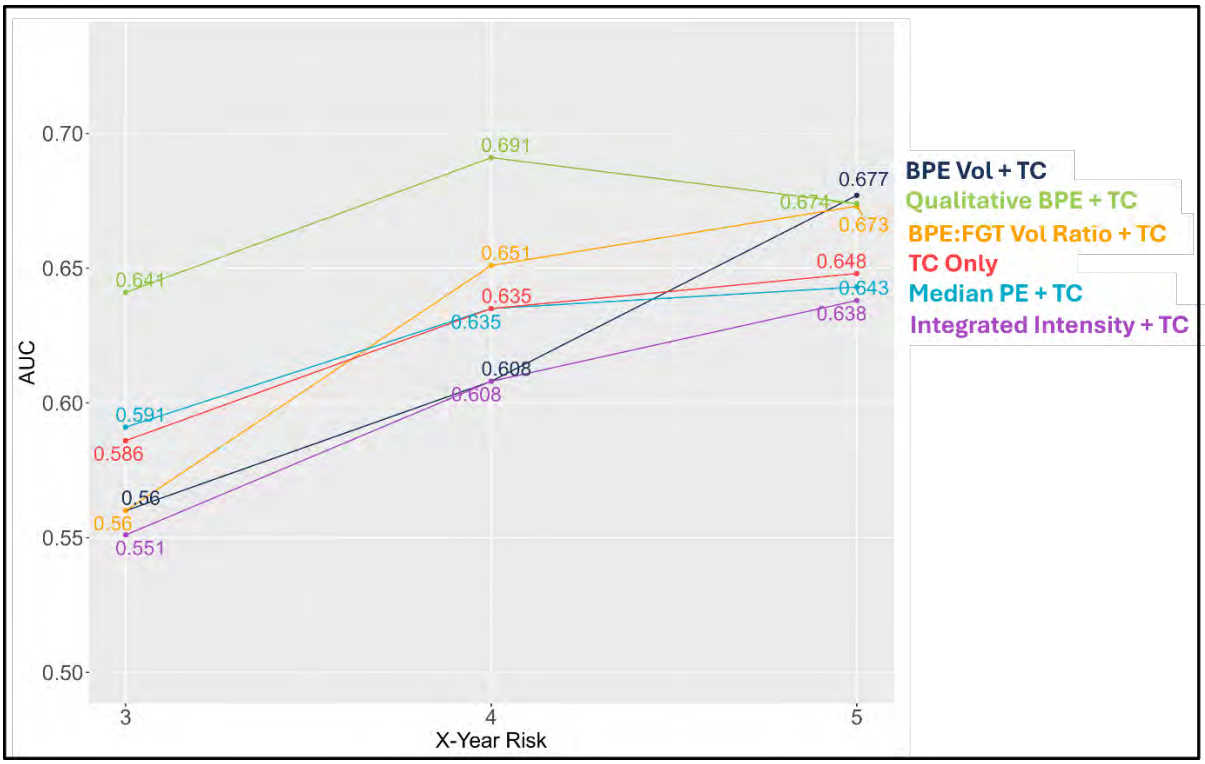

Figure 6.3. Performance across 3-year, 4-year and 5-year cancer risk prediction.

6.6 Conclusion for multimodal combination of imaging and clinical risk factors

In this chapter, I report findings that support the hypothesis of multimodal risk prediction outperforming univariate TC Only prediction. Through exploring various cohort subset strategies, I also show that quantitative BPE + TC outperforms qualitative BPE + TC in some cases and outperforms TC Only in most cases. My work contributes to the effort to show that multimodal approaches can better represent an individual's biology and provide incremental improvement of risk prediction over using clinical risk score alone. With this cohort, multimodal models showed numerical advantage in AUCs over univariate ones, albeit not overwhelming improvements in prediction accuracy. With that in mind, we are well positioned to further discuss the limitations of our study and explore future directions both for the project and the field.

Chapter 7
Conclusion and Future Work

This chapter will begin by addressing the limitations of our work, in Section 6.1. In Section 7.2, I discuss the next steps for this dataset, separating them into short-term and long-term goals. In Section 7.3, I offer some opinions on what may be next for similar work in the field that is also within the context of multimodal risk prediction. Lastly, in Section 7.4, I conclude with the contributions that this dissertation makes towards clinical and biomedical informatics research.

7.1 Limitations of this dissertation work

There were several limitations for our main work in Chapters 5 and 6. Our study cohort involved women who are at high risk of developing cancer, thus limiting the generalizability of our findings to the larger population of average-risk women. Our research is limited to MRI markers and does not directly apply to mammography. There are clinical trials such as the EA1141 (Kuhl et al., 2023) and DENSE Trial (Bakker et al., 2019) studies which perform supplementary MR screening for women who have mammographically dense breasts, which may include average-risk women. However, these studies investigated cancer detection instead of predicting future cancer, thus serving different purposes from our work. Future work predicting cancer risk in average-risk women would have to leverage imaging markers derived from mammographic imaging since that is the conventional screening modality for the average-risk population.

Our work could also have been strengthened by garnering a larger study cohort. As mentioned in Chapter 4, choosing a larger sampling percentage would benefit statistical power, but also demands more time and effort to collect clinical risk factors and derive quantitative metrics for more patients. One way to ensure a larger sample size would be to begin with a multi-institutional dataset which would introduce more cancer positive cases. This approach comes with its own challenges and opportunities, which I will discuss in Section 7.2.2 as part of long-term next steps. Another factor that also affected our final study cohort size was the loss of patients due to technical failures in the intermediate steps of our image processing pipeline; roughly estimating about ten additional positive cases and also ten more controls. The failures were often caused by variabilities in image quality or file conversion for older exams. The loss of patients due to processing failures may be remedied by either future attempts at salvaging images

using specialized open-source tools or finding replacement patients who also satisfy the eligibility criteria as those who are lost. It is worth noting that this issue is not unique to our work; all studies leveraging medical image data acquired over many years would have to contend with the challenge of variable data quality.

As pointed out in Chapter 5, our MR images were acquired on multiple scanner systems and field strengths. Differences in scanner magnetic field strength may lead to differences in image intensity values especially in the case of pre- and post-contrast changes in enhancement. This ultimately affects the magnitude of signal change which we calculate as part of quantification of BPE and influences the selection of enhancement thresholds. In Chapter 5, rather than selecting a single threshold for enhancement, we derived optimized thresholds for each field strength during every fold of the model training phase. Our approach is intended to reduce the influence of field strength differences on threshold selection of image markers and ultimately, marker prediction performance. There is a possible limitation in that our approach may depend on the training fold sample size and therefore would need to be validated in different datasets. Alternatives to our approach of operating with imaging marker values would involve image normalization methods such as RAVEL (Fortin et al., 2016) or ComBat (Fortin et al., 2017), seen in other multi-scanner studies. Future work that involves images from multiple field strengths would be strengthened by comparing various image data harmonization approaches.

Our study is also limited by the missingness of specific clinical risk factors. This missingness may stem from the inherent lack of data availability such as low rates of BRCA testing performed on even high-risk women. Our findings in Section 6.3.1 illustrated that the presence of BRCA1 or BRCA2 mutation greatly increases an individual's TC risk score. If BRCA test results are hypothetically present for all individuals in our study cohort, we would be assured that every person's TC risk score is calculated using the same availability of that risk factor. Considering every individual has different clinical necessities and healthcare coverage, providers likely make different care decisions for each person regarding testing and intervention, which includes whether to offer BRCA testing. For the specific example of BRCA testing result availability, it is perhaps acceptable to acknowledge that some clinical risk factors are unavailable since they were not necessary to capture as part of clinical care. Therefore, when using risk assessment tools, one should not assume that risk factor completeness is uniform across individuals. The performance of risk assessment tools is then also inherently influenced by the availability of clinical data; a risk score calculated using only age as input will not perform the same as a risk score calculated using multiple risk factors. Another source of data missingness may also stem from human error and the limitations in my approach to chart reviews, potentially missing some risk factors that were indeed present in the medical records. For every clinical risk factor, we are limited by the quality of relevant key search terms used, as well as the navigational proficiency to find relevant notes or reports that hold the desired information. While every effort was made to familiarize myself with the medical records environment and work with research staff who have experience with record lookups, it is fair to acknowledge that the lookups may not have been perfectly comprehensive. A potential way to improve data collection would perhaps involve soliciting guidance from experienced providers routinely involved in using such risk

assessment tools as to the best methods to find specific risk factors in the medical record. Once the same progress notes have been identified, consistency of information extraction can be evaluated by using multiple readers to conduct lookup and assess kappa coefficients to compare how the readers differ in their assessment of the notes.

7.2 Next steps for this work

7.2.1 Short-term next steps

Now we will offer several short-term next steps that are meant to be feasible in the near future. They are roughly divided into those that involve expanding the current study cohort sample, and those that do not require that expansion.

Without changing the study cohort, one can explore different ways to refine the image processing. Imaging data harmonization approaches such as RAVEL (Fortin et al., 2016) or ComBat (Fortin et al., 2017) can address image intensity differences that may be caused by multiple scanner systems as well as different field strengths. This would likely impact quantitative BPE markers and by extension, their performance in predicting 5-year cancer. This would also potentially result in different optimal enhancement thresholds for each of the quantitative BPE markers. This aligns with the fact that there is currently no universally agreed approach to set an enhancement threshold (Lam et al., 2019; Niell et al., 2021). While harmonizing data at the image level may not entirely eliminate the issue of variation in optimal thresholds, it would offer a more standardized way of harmonizing data that is less susceptible to sampling biases of any specific dataset.

Alternatively, one can also investigate using different image data to derive quantitative BPE through selecting a later post-contrast image acquisition in the dynamic contrast-enhanced sequence such as in Niell et. al. instead of using the first post-contrast phase as we have done. While Niell et. al. (Niell et al., 2021) did not find significant differences in predictive performance of quantitative markers derived from different post-contrast phases, this aspect remains underexplored and it is still unknown whether there is a particular post-contrast phase that is best for generating predictive quantitative markers.

Another additional step for statistical analysis would be to apply a more complex analysis approach using time-dependent risk modeling such as Cox regression. This would provide the proportional hazards of each marker's association with new cancer. This is possible to do since the date of diagnosis for positives and date of last follow-up for controls are available for our study cohort. The current analysis approach simply predicts the presence or absence of a cancer diagnosis within 5 years, but does not account for the time-to-event aspect in the prediction using each marker. This would be an interesting short-term step for this project.

Apart from the core clinical risk factors we selected for our study, future work could explore adding other clinical risk factors pertaining to the individual's biology. Specifically,

this would include the woman's age at first live birth of an offspring, and hormone-related therapies either for cancer risk reduction or for management of menopausal symptoms. Since the age at first live birth is incorporated into Tyrer-Cuzick, Gail and BCSC models (Cintolo-Gonzalez et al., 2017a), obtaining this information would enable us to produce risk scores using these models in the future. We did not have the opportunity to include these risk factors in our lookup because of the significant additional time it would involve, therefore making them appropriate goals to explore in the long-term. Information on an individual's previous pregnancies and deliveries are recorded in progress notes that are not often associated with breast cancer screening, so finding them for our entire study cohort would entail a non-trivial effort. History of personal use of hormone-related therapy will also be difficult to collect, especially since it would involve information such as dosage and duration, which are anecdotally found to have sporadic completeness. Age at first live birth and use of hormone-related therapies are risk factors that provide information for an individual's hormonal exposure, influencing both cancer risk score and fibroglandular tissue biology. These risk factors may also have a moderately high degree of missingness while expected to offer a limited amount of improvement in prediction performance, therefore not prioritized during our risk factor lookup.

7.2.2 Long-term next steps

Looking further ahead, there are several other possible directions to extend the research in this dissertation. I classify these as long-term goals due to the significant amount of work that would be involved, as well as natural extensions that answer different questions from what we have done.

There are also opportunities that may be taken in the aspect of the study period. Through continued follow-up, we can obtain cancer outcomes for more screening patients, thereby increasing the pool of patients eligible to be added to our study cohort. This would entail risk factor lookup and quantitative marker derivation for every new patient but would strengthen the study from a sample size aspect. This may potentially improve predictive performance of both clinical risk models as well as quantitative markers by virtue of data recency. With regards to clinical data, newer records are in theory easier to find and collect, potentially leading to improved data completeness. With regards to imaging data, newer acquisitions are likely to be more standardized in field strength and acquisition parameters, potentially having less heterogeneity in data quality. Another added benefit from having a longer study period would be the opportunity to explore year-by-year predictions for more years, enabling us to observe whether different combinations of variables are more predictive at different time scales. As a woman's BPE may experience fluctuations around the time of menopause (King et al., 2012; Okuma et al., 2024), the predictive performance of enhancement-derived markers may also change accordingly, in addition to increasing risk of developing cancer with age.

Another future possibility would be to involve imaging sequences beyond DCE, such as diffusion-weighted imaging. Diffusion-weighted MRI markers have been found to

detect malignancies as a screening tool (Amornsiripanitch et al., 2019), differentiate malignant and benign lesions (Partridge et al., 2010), and predict neoadjuvant treatment response (Partridge et al., 2018), but there is still valuable opportunity in assessing their ability to predict cancer development. Diffusion-weighted MRI provides insights into the tissue microstructure, thus supplying information that is different from DCE and potentially useful for early detection of changes in cellular environment prior to cancer development. While the main results of this dissertation in Chapters 5 and 6 leveraged DCE images for quantitative BPE metrics, our publication (Surento et al., 2025) described in Section 3.1 involved markers derived from diffusion-weighted imaging. We did not find that diffusion-weighted imaging measures such as BPS and ADC metrics to be associated significantly with TC risk group, while considering that we were limited by a modest sample size of 77 individuals and lack of 5-year cancer outcomes. Performing a similar multiparametric investigation of cancer risk for the full study cohort would be of high scientific interest, but also would require substantial additional effort to process diffusion-weighted image data and perform quality control of segmentation masks and derived measures, making this a long-term goal.

Another potential long-term next step for this project would be to validate the findings by using multi-institutional data. The current dissertation work utilized data from a single institution, and a future iteration could be strengthened by including data from more than one institution from the beginning. The same exclusion criteria and case-cohort sampling may potentially be applied the same way but starting with a much larger pool of patients. Besides a boost in sample size, the findings from a multi-institutional study will likely be more robust and generalizable compared to a single-institutional study. The challenges of a multi-institutional study are not trivial and would include but not be limited to: differences in clinical data quality between the institutions, differences in image data acquisition parameters, a larger volume of work and effort needed to extract imaging markers and perform clinical risk factor lookup, larger computational and storage needs to accommodate the larger volume of data, and the creation of safe data handling principles to protect patient privacy amidst cross-institutional data sharing.

While this dissertation has focused on the prediction of first cancer within 5 years of the screening MRI, another direction to explore would be to predict recurrent cancer within some time after a specific screening MRI. This would represent an entirely different topic of investigation, as the presence of a prior cancer diagnosis may nullify the validity of including scores from risk assessment models designed to predict first breast cancer. It would likely involve different cohort selection criteria and data collection options. The same clinical risk factors may perhaps still be relevant, just not for the purpose of calculating a risk score using a risk assessment model. Predicting recurrent cancer will probably involve markers extracted from pre- and post-treatment image data, alongside new risk factors involving lesion characteristics and therapy information such as duration and dose. An obvious image processing challenge would involve the usage of post-operative breast scans that show missing tissue after surgical resection. Downstream

processing steps such as image segmentation and marker quantitation workflows would need to be trained to recognize the resected regions as abnormal and avoid mislabeling them as tumors. Similar to predicting first cancer, the motivation of predicting risk of cancer recurrence would also be for tailoring surveillance strategies and improved planning of intervention strategies such as prescription of risk-reducing medication or surgical interventions. While beyond the scope of this dissertation, prediction of recurrent cancers could be a valuable new direction for investigation.

7.3 Next steps for the field

Building on what we have learned from our dissertation work, we would like to offer several possible directions for future studies in the field of multimodal risk prediction of disease.

There have not been many studies that leverage screening MRI to predict risk of cancer development. In the most relevant study to date, Portnoi et al. developed an MRI-based image-only model which used a 2D maximum intensity projection (MIP) images, combining the highest-intensity voxel of every slice into one, to predict a binary outcome of cancer diagnosis or absence status within 5 years, comparing its performance with Tyrer-Cuzick and a logistic regression model incorporating traditional clinical risk factors and qualitative BPE (Portnoi et al., 2019). Training and testing were performed with 10-fold cross validation on 1656 MR images from 1183 unique patients. They found that the image-only deep learning model (AUC: 0.638) outperformed both the risk-factor-logistic-regression (AUC: 0.558) and Tyrer Cuzick (AUC: 0.493) models. Strengths of this study include the substantial training dataset size of MR images, practical handling of converting 3D MR images to 2D MIP projections, and the result demonstrating image-only is capable of predicting cancer risk well. While this study did not combine clinical risk factors and medical images, it serves as an example of using MRI instead of mammograms for cancer prediction. The three-dimensionality of MR images presents both opportunities and challenges; it preserves the rich local-level imaging features for models to learn from, while model training that involves 3D data will also be more computationally expensive than models learning from 2D data. Notwithstanding, an ideal direction for the field would be to develop a multimodal cancer risk prediction model which leverages a combination of 3D MRI and clinical risk factors. Based on the findings in my dissertation, it is my expectation that such an approach would outperform current models that collapse 3D images into 2D.

An important precursor of multimodal work would be a better integration between the disparate databases that store imaging and non-imaging clinical data. While this is not advocating to change the current configuration of keeping imaging and other clinical data separate, it would be a call to an intermediate portal or environment that is connected to both and allows for a quick query for both data types. This would empower a multimodal risk prediction tool to be used during a clinic visit, allowing for shared decision-making

between provider and patient regarding personal cancer risk. The challenge would be to have a computational infrastructure that allows for on-demand running inference on imaging and clinical data. Alternatively, even if it is not real-time, such multimodal risk inference may be performed before the clinic visit and discussed later.

Another important future direction for the field, which may arguably be relevant for any new advances, is the need for external validation of predictive model performance. Regardless of what form of future prediction is involved, findings discovered using single-site institutional data will always need to have their robustness tested by applying to an external dataset. This ensures that the inferences of such findings are reproducible using other datasets and not simply an artifact of the inherent biases of regional population sampling. One of the significant learnings from the work done in this dissertation is the reality of real-world clinical data, rife with variable quality and missingness. Any new predictive model needs to be able to withstand the test of heterogeneity and incompleteness for it to be robust to be deployed on new prospective data which may harbor new variations and challenges.

While beyond the scope of this dissertation work, an impactful future direction for cancer risk prediction modeling would be to develop a clinical decision support tool. The example alluded to in our dissertation work would be for better stratification of women who are considered to be at high risk of cancer; enabling consideration of de-escalation or escalation of screening and intervention strategies. Implementation of a clinical decision support tool certainly needs to involve stakeholders in the design, development and deployment phases. One of the important technical considerations would revolve around defining acceptable prediction performance for the tool to have clinical utility. For instance, as summarized in Table 6.2, selecting specific cutoffs for specificity will result in a corresponding sensitivity and vice versa, prompting conversations around the tradeoff of balancing high sensitivity and high specificity to differing extents. In a cancer screening scenario, this would involve weighing the consequences of false negatives where cancers are missed, or false positives where additional procedures may be prescribed but not truly needed..

7.4 Contributions of this dissertation

My contributions can be viewed from two different perspectives: clinical and biomedical informatics.

From the clinical perspective, conventional risk assessment models are standard-of-care in breast cancer screening. The first clinical contribution of the work in this dissertation is improving the prediction performance of a conventional risk assessment model by incorporating imaging markers. Conventional risk assessment markers leverage risk factors that may be commonly used to group patients on a population level,

while imaging markers offer a direct reflection of the individual's biology at a specific point in time. There is yet potential to involve other imaging markers from different modalities, or variations on quantitation using DCE images commonly used in breast MRI screening.

The second clinical contribution of our work is making a step towards personalizing care for women already considered to be at high risk of cancer. A future iteration of a multimodal risk prediction model may be used to provide risk information to be used by physicians in better taioring screening and intervention strategies. There is opportunity to identify individuals at the lower end of the high-risk spectrum and offer a de-escalation of screening perhaps by decreasing the screening frequency from annual to biannual. For individuals identified as being at the higher end of the high-risk spectrum, escalation options might be considered, such as prescription of risk-reducing aromatase inhibitors or selective estrogen receptor modulators (SERMs).

From a biomedical informatics perspective, our work has demonstrated the value of using quantitative imaging markers in the use of cancer risk prediction. Improving the prediction performance of a conventional risk assessment tool implies that there may be subtle differences in the screening MRIs of individuals who will eventually develop cancer and those who will not. This potentially opens the possibility of developing other quantitative imaging markers that capture those differences better, which may be used as more impactful markers to predict cancer risk and stratify patients.

The second contribution towards informatics work is the construction of a clean, reusable dataset of clinical risk factors that may be leveraged in other studies. We identified core risk factors that may significantly affect an individual's risk and may be reused to generate risk scores using other risk assessment models. Furthermore, the database design of which risk factors to collect may also serve as a guide for future work aimed at risk prediction. While the exact risk factors may be different for applications in predicting other diseases, principles surrounding the need to use risk factors adjacent to the time of imaging may be important and transferable. This is especially relevant for the intent of developing a risk prediction model that can be run prospectively in a new patient whose medical information may be limited.

The central hypothesis of this dissertation is that combining imaging and clinical risk factors will produce better prediction performance than using either alone. We have shown that imaging markers can improve the performance of a conventional risk assessment tool in predicting cancer risk.

References

Amir, E., Evans, D. G., Shenton, A., Lalloo, F., Moran, A., Boggis, C., Wilson, M., & Howell, A. (2003). Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *Journal of Medical Genetics*, *40*(11), 807–814. https://doi.org/10.1136/jmg.40.11.807

Amornsiripanitch, N., Bickelhaupt, S., Shin, H. J., Dang, M., Rahbar, H., Pinker, K., & Partridge, S. C. (2019). Diffusion-weighted MRI for Unenhanced Breast Cancer Screening. *Radiology*, *293*(3), 504–520. https://doi.org/10.1148/radiol.2019182789

Arasu, V. A., Miglioretti, D. L., Sprague, B. L., Alsheik, N. H., Buist, D. S. M., Henderson, L. M., Herschorn, S. D., Lee, J. M., Onega, T., Rauscher, G. H., Wernli, K. J., Lehman, C. D., & Kerlikowske, K. (2019). Population-Based Assessment of the Association Between Magnetic Resonance Imaging Background Parenchymal Enhancement and Future Primary Breast Cancer Risk. *Journal of Clinical Oncology*, *37*(12), 954–963. https://doi.org/10.1200/JCO.18.00378

Arefan, D., Mohamed, A. A., Berg, W. A., Zuley, M. L., Sumkin, J. H., & Wu, S. (2020). Deep learning modeling using normal mammograms for predicting breast cancer risk. *Medical Physics*, *47*(1), 110–118. https://doi.org/10.1002/mp.13886

Arian, A., Dinas, K., Pratilas, G. C., & Alipour, S. (2022). The Breast Imaging-Reporting and Data System (BI-RADS) Made Easy. *Iranian Journal of Radiology*, *19*(1). https://doi.org/10.5812/iranjradiol-121155

Bakker, M. F., De Lange, S. V., Pijnappel, R. M., Mann, R. M., Peeters, P. H. M., Monninkhof, E. M., Emaus, M. J., Loo, C. E., Bisschops, R. H. C., Lobbes, M. B. I., De Jong, M. D. F., Duvivier, K. M., Veltman, J., Karssemeijer, N., De Koning, H. J., Van Diest, P. J., Mali, W. P. T. M., Van Den Bosch, M. A. A. J., Veldhuis, W. B., & Van Gils, C. H. (2019). Supplemental MRI Screening for Women

with Extremely Dense Breast Tissue. *New England Journal of Medicine*, *381*(22), 2091–2102. https://doi.org/10.1056/NEJMoa1903986

Barlow, W. E., White, E., Ballard-Barbash, R., Vacek, P. M., Titus-Ernstoff, L., Carney, P. A., Tice, J. A., Buist, D. S. M., Geller, B. M., Rosenberg, R., Yankaskas, B. C., & Kerlikowske, K. (2006). Prospective Breast Cancer Risk Prediction Model for Women Undergoing Screening Mammography. *JNCI: Journal of the National Cancer Institute*, *98*(17), 1204–1214. https://doi.org/10.1093/jnci/djj331

Berg, W. A., Blume, J. D., Cormack, J. B., Mendelson, E. B., & Lehrer, D. (n.d.). *Combined Screening With Ultrasound and Mammography vs Mammography Alone in Women at Elevated Risk of Breast Cancer*.

Berg, W. A., Zhang, Z., Lehrer, D., Jong, R. A., Pisano, E. D., Barr, R. G., Böhm-Vélez, M., Mahoney, M. C., Evans, W. P., Larsen, L. H., Morton, M. J., Mendelson, E. B., Farria, D. M., Cormack, J. B., Marques, H. S., Adams, A., Yeh, N. M., & Gabrielli, G. (2012). Detection of Breast Cancer With Addition of Annual Screening Ultrasound or a Single Screening MRI to Mammography in Women With Elevated Breast Cancer Risk. *JAMA*, *307*(13), 1394. https://doi.org/10.1001/jama.2012.388

Bhatelia, K., Singh, K., & Singh, R. (2014). TLRs: Linking inflammation and breast cancer. *Cellular Signalling*, *26*(11), 2350–2357. https://doi.org/10.1016/j.cellsig.2014.07.035

Boyd, N. F., Sun, L., Stone, J., Fishell, E., Jong, R. A., & Chiarelli, A. (2007). Mammographic Density and the Risk and Detection of Breast Cancer. *The New England Journal of Medicine*.

Brentnall, A. R., Harkness, E. F., Astley, S. M., Donnelly, L. S., Stavrinos, P., Sampson, S., Fox, L., Sergeant, J. C., Harvie, M. N., Wilson, M., Beetles, U., Gadde, S., Lim, Y., Jain, A., Bundred, S., Barr, N., Reece, V., Howell, A., Cuzick, J., & Evans, D. G. R. (2015). Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Research*, *17*(1), 147. https://doi.org/10.1186/s13058-015-0653-5

Brodersen, J., & Siersma, V. D. (2013). Long-Term Psychosocial Consequences of False-Positive Screening Mammography. *The Annals of Family Medicine*, *11*(2), 106–115. https://doi.org/10.1370/afm.1466

Byng, J. W., Yaffe, M. J., Jong, R. A., Shumak, R. S., Lockwood, G. A., Tritchler, D. L., & Boyd, N. F. (1998). Analysis of Mammographic Density and Breast Cancer Risk from Digitized Mammograms. *RadioGraphics*, *32*, 1587–1898.

Cai, J., & Zeng, D. (2004). Sample Size/Power Calculation for Case–Cohort Studies. *Biometrics*, *60*(4), 1015–1024. https://doi.org/10.1111/j.0006-341X.2004.00257.x

Cai, J., & Zeng, D. (2007). Power Calculation for Case–Cohort Studies with Nonrare Events. *Biometrics*, *63*(4), 1288–1295. https://doi.org/10.1111/j.1541-0420.2007.00838.x

Castañeda-Gill, J., & Vishwanatha, J. (2016). Antiangiogenic mechanisms and factors in breast cancer treatment. *Journal of Carcinogenesis*, *15*(1), 1. https://doi.org/10.4103/1477-3163.176223

Caswell-Jin, J. L., Sun, L. P., Munoz, D., Lu, Y., Li, Y., Huang, H., Hampton, J. M., Song, J., Jayasekera, J., Schechter, C., Alagoz, O., Stout, N. K., Trentham-Dietz, A., Lee, S. J., Huang, X., Mandelblatt, J. S., Berry, D. A., Kurian, A. W., & Plevritis, S. K. (2024). Analysis of Breast Cancer Mortality in the US—1975 to 2019. *JAMA*, *331*(3), 233. https://doi.org/10.1001/jama.2023.25881

Chen, J., Pee, D., Ayyagari, R., Graubard, B., Schairer, C., Byrne, C., Benichou, J., & Gail, M. H. (2006). Projecting Absolute Invasive Breast Cancer Risk in White Women With a Model That Includes Mammographic Density. *JNCI: Journal of the National Cancer Institute*, *98*(17), 1215–1226. https://doi.org/10.1093/jnci/djj332

Cheung, Y.-C., Lin, Y.-C., Wan, Y.-L., Yeow, K.-M., Huang, P.-C., Lo, Y.-F., Tsai, H.-P., Ueng, S.-H., & Chang, C.-J. (2014). Diagnostic performance of dual-energy contrast-enhanced subtracted mammography in dense breasts compared to mammography alone: Interobserver blind-reading analysis. *European Radiology*, *24*(10), 2394–2403. https://doi.org/10.1007/s00330-014-3271-1

Choi, E. J., Choi, H., Choi, S. A., & Youk, J. H. (2016). Dynamic contrast-enhanced breast magnetic resonance imaging for the prediction of early and late recurrences in breast cancer. *Medicine*, *95*(48), e5330. https://doi.org/10.1097/MD.0000000000005330

Cintolo-Gonzalez, J. A., Braun, D., Blackford, A. L., Mazzola, E., Acar, A., Plichta, J. K., Griffin, M., & Hughes, K. S. (2017a). Breast cancer risk models: A comprehensive overview of existing models, validation, and clinical applications. *Breast Cancer Research and Treatment*, *164*(2), 263–284. https://doi.org/10.1007/s10549-017-4247-z

Cintolo-Gonzalez, J. A., Braun, D., Blackford, A. L., Mazzola, E., Acar, A., Plichta, J. K., Griffin, M., & Hughes, K. S. (2017b). Breast cancer risk models: A comprehensive overview of existing models, validation, and clinical applications. *Breast Cancer Research and Treatment*, *164*(2), 263–284. https://doi.org/10.1007/s10549-017-4247-z

Claus, E. B., Risch, N., & Thompson, W. D. (1994). Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*, *73*(3), 643–651. https://doi.org/10.1002/1097-0142(19940201)73:3%3C643::AID-CNCR2820730323%3E3.0.CO;2-5

Couch, F. J., DeShano, M. L., Blackwood, M. A., Calzone, K., Stopfer, J., Campeau, L., Ganguly, A., Rebbeck, T., Weber, B. L., Jablon, L., Cobleigh, M. A., Hoskins, K., & Garber, J. E. (1997). *BRCA1* Mutations in Women Attending Clinics That Evaluate the Risk of Breast Cancer. *New England Journal of Medicine*, *336*(20), 1409–1415. https://doi.org/10.1056/NEJM199705153362002

Dontchos, B. N., Rahbar, H., Partridge, S. C., Korde, L. A., Lam, D. L., Scheel, J. R., Peacock, S., & Lehman, C. D. (2015). Are Qualitative Assessments of Background Parenchymal Enhancement, Amount of Fibroglandular Tissue on MR Images, and Mammographic Density Associated with Breast Cancer Risk? *Radiology*, *276*(2), 371–380. https://doi.org/10.1148/radiol.2015142304

D'Orsi, C. J., Sickles, E. A., Mendelson, E. B., & Morris, E. A. (2013). *2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System* (5th ed.). American College of Radiology. https://books.google.com/books?vid=ISBN155903016X

Duggan, C., Trapani, D., Ilbawi, A. M., Fidarova, E., Laversanne, M., Curigliano, G., Bray, F., & Anderson, B. O. (2021). National health system characteristics, breast cancer stage at diagnosis, and breast cancer mortality: A population-based analysis. *The Lancet Oncology*, *22*(11), 1632–1642. https://doi.org/10.1016/S1470-2045(21)00462-9

Elmore, J. G., Barton, M. B., Moceri, V. M., Polk, S., Arena, P. J., & Fletcher, S. W. (1998). Ten-Year Risk of False Positive Screening Mammograms and Clinical Breast Examinations. *New England Journal of Medicine*, *338*(16), 1089–1096. https://doi.org/10.1056/NEJM199804163381601

Fakhrejahani, E., & Toi, M. (2014). Antiangiogenesis Therapy for Breast Cancer: An Update and Perspectives from Clinical Trials. *Japanese Journal of Clinical Oncology*, *44*(3), 197–207. https://doi.org/10.1093/jjco/hyt201

Fallenberg, E. M., Schmitzberger, F. F., Amer, H., Ingold-Heppner, B., Balleyguier, C., Diekmann, F., Engelken, F., Mann, R. M., Renz, D. M., Bick, U., Hamm, B., & Dromain, C. (2017). Contrast-enhanced spectral mammography vs. Mammography and MRI – clinical performance in a multi-reader evaluation. *European Radiology*, *27*(7), 2752–2764. https://doi.org/10.1007/s00330-016-4650-6

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., & Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*, 149–170. https://doi.org/10.1016/j.neuroimage.2017.08.047

Fortin, J.-P., Sweeney, E. M., Muschelli, J., Crainiceanu, C. M., & Shinohara, R. T. (2016). Removing inter-subject technical variability in magnetic resonance imaging studies. *NeuroImage*, *132*, 198–212. https://doi.org/10.1016/j.neuroimage.2016.02.036

Frank, T. S., Manley, S. A., Olopade, O. I., Cummings, S., Garber, J. E., Bernhardt, B., Antman, K., Russo, D., Wood, M. E., Mullineau, L., Isaacs, C., Peshkin, B., Buys, S., Venne, V., Rowley, P. T., Loader, S., Offit, K., Robson, M., Hampel, H., … Thomas, A. (1998). Sequence analysis of BRCA1 and BRCA2: Correlation of mutations with family history and ovarian cancer risk. *Journal of Clinical Oncology*, *16*(7), 2417–2425. https://doi.org/10.1200/JCO.1998.16.7.2417

Fuller, A. M., Olsson, L. T., Midkiff, B. R., Kirk, E. L., McNaughton, K. K., Calhoun, B. C., & Troester, M. A. (2019). Vascular density of histologically benign breast tissue from women with breast cancer: Associations with tissue composition and tumor characteristics. *Human Pathology*, *91*, 43–51. https://doi.org/10.1016/j.humpath.2019.06.003

Gail, M. H. (2015). Twenty-five Years of Breast Cancer Risk Models and Their Applications. *JNCI Journal of the National Cancer Institute*, *107*(5), djv042–djv042. https://doi.org/10.1093/jnci/djv042

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., & Mulvihill, J. J. (1989). Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. *JNCI Journal of the National Cancer Institute*, *81*(24), 1879–1886. https://doi.org/10.1093/jnci/81.24.1879

Garrison, L. P., Babigumira, J. B., Masaquel, A., Wang, B. C. M., Lalla, D., & Brammer, M. (2015). The Lifetime Economic Burden of Inaccurate HER2 Testing: Estimating the Costs of False-Positive and False-Negative HER2 Test Results in US Patients with Early-Stage Breast Cancer. *Value in Health*, *18*(4), 541–546. https://doi.org/10.1016/j.jval.2015.01.012

Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., Jemal, A., & Siegel, R. L. (2022). Breast Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(6), 524–541. https://doi.org/10.3322/caac.21754

Giuliano, V., & Giuliano, C. (2013). Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. *Clinical Imaging*, *37*(3), 480–486. https://doi.org/10.1016/j.clinimag.2012.09.018

Grubstein, A., Rapson, Y., Benzaquen, O., Rozenblatt, S., Gadiel, I., Atar, E., Yerushalmi, R., & Cohen, M. J. (2018). Comparison of background parenchymal enhancement and fibroglandular density at breast magnetic resonance imaging between BRCA gene mutation carriers and non-carriers. *Clinical Imaging*, *51*, 347–351. https://doi.org/10.1016/j.clinimag.2018.06.010

Hahn, S. Y., Ko, E. S., Han, B.-K., Lim, Y., Gu, S., & Ko, E. Y. (2016). Analysis of factors influencing the degree of detectability on diffusion-weighted MRI and diffusion background signals in patients with invasive breast cancer. *Medicine*, *95*(27), e4086. https://doi.org/10.1097/MD.0000000000004086

Hoerger, T. J., Ekwueme, D. U., Miller, J. W., Uzunangelov, V., Hall, I. J., Segel, J., Royalty, J., Gardner, J. G., Smith, J. L., & Li, C. (2011). Estimated Effects of the National Breast and Cervical Cancer Early Detection Program on Breast Cancer Mortality. *American Journal of Preventive Medicine*, *40*(4), 397–404. https://doi.org/10.1016/j.amepre.2010.12.017

Hooley, R. J., Greenberg, K. L., Stackhouse, R. M., Geisel, J. L., Butler, R. S., & Philpotts, L. E. (2012). Screening US in Patients with Mammographically Dense Breasts: Initial Experience with Connecticut Public Act 09-41. *Radiology*, *265*(1), 59–69. https://doi.org/10.1148/radiol.12120621

Houssami, N., Ciatto, S., Martinelli, F., Bonardi, R., & Duffy, S. W. (2009). Early detection of second breast cancers improves prognosis in breast cancer survivors. *Annals of Oncology*, *20*(9), 1505–1510. https://doi.org/10.1093/annonc/mdp037

Hu, X., Jiang, L., Li, Q., & Gu, Y. (2017). Quantitative assessment of background parenchymal enhancement in breast magnetic resonance images predicts the risk of breast cancer. *Oncotarget*, *8*(6), 10620–10627. https://doi.org/10.18632/oncotarget.13538

Hu, X., Jiang, L., You, C., & Gu, Y. (2021). Fibroglandular Tissue and Background Parenchymal Enhancement on Breast MR Imaging Correlates With Breast Cancer. *Frontiers in Oncology*, *11*, 616716. https://doi.org/10.3389/fonc.2021.616716

Jacobi, C. E., De Bock, G. H., Siegerink, B., & Van Asperen, C. J. (2009). Differences and similarities in breast cancer risk assessment models in clinical practice: Which model to choose? *Breast Cancer Research and Treatment*, *115*(2), 381–390. https://doi.org/10.1007/s10549-008-0070-x

Jesinger, R. A., Lattin, G. E., Ballard, E. A., Zelasko, S. M., & Glassman, L. M. (2011). Vascular Abnormalities of the Breast: Arterial and Venous Disorders, Vascular Masses, and Mimic Lesions with Radiologic-Pathologic Correlation. *RadioGraphics*, *31*(7), E117–E136. https://doi.org/10.1148/rg.317115503

Jochelson, M. S., & Lobbes, M. B. I. (2021). Contrast-enhanced Mammography: State of the Art. *Radiology*, *299*(1), 36–48. https://doi.org/10.1148/radiol.2021201948

Johns, P. C., & Yaffe, M. J. (1987). X-ray characterisation of normal and neoplastic breast tissues. *Physics in Medicine and Biology*, *32*(6), 675–695. https://doi.org/10.1088/0031-9155/32/6/002

Keller, B. M., Chen, J., Daye, D., Conant, E. F., & Kontos, D. (2015). Preliminary evaluation of the publicly available Laboratory for Breast Radiodensity Assessment (LIBRA) software tool: Comparison of fully automated area and volumetric density measures in a case–control study with digital

mammography. *Breast Cancer Research*, *17*(1), 117. https://doi.org/10.1186/s13058-015-0626-8

Kim, J. Y., Kim, S. H., Kim, Y. J., Kang, B. J., An, Y. Y., Lee, A. W., Song, B. J., Park, Y. S., & Lee, H. B. (2015). Enhancement parameters on dynamic contrast enhanced breast MRI: Do they correlate with prognostic factors and subtypes of breast cancers? *Magnetic Resonance Imaging*, *33*(1), 72–80. https://doi.org/10.1016/j.mri.2014.08.034

King, V., Brooks, J. D., Bernstein, J. L., Reiner, A. S., Pike, M. C., & Morris, E. A. (2011). Background Parenchymal Enhancement at Breast MR Imaging and Breast Cancer Risk. *Radiology*, *260*(1), 50–60. https://doi.org/10.1148/radiol.11102156

King, V., Gu, Y., Kaplan, J. B., Brooks, J. D., Pike, M. C., & Morris, E. A. (2012). Impact of menopausal status on background parenchymal enhancement and fibroglandular tissue on breast MRI. *European Radiology*, *22*(12), 2641–2647. https://doi.org/10.1007/s00330-012-2553-8

Kriege, M., Brekelmans, C. T. M., Zonderland, H. M., Kok, T., & Meijer, S. (2004). Efficacy of MRI and Mammography for Breast-Cancer Screening in Women with a Familial or Genetic Predisposition. *N Engl j Med*.

Kristensen, T., Knutsson, M., Wehland, M., Laursen, B., Grimm, D., Warnke, E., & Magnusson, N. (2014). Anti-Vascular Endothelial Growth Factor Therapy in Breast Cancer. *International Journal of Molecular Sciences*, *15*(12), 23024–23041. https://doi.org/10.3390/ijms151223024

Kuhl, C. K., Bieling, H. B., Gieseke, J., Kreft, B. P., Sommer, T., Lutterbey, G., & Schild, H. H. (1997). Healthy premenopausal breast parenchyma in dynamic contrast-enhanced MR imaging of the breast: Normal contrast medium enhancement and cyclical-phase dependency. *Radiology*, *203*(1), 137–144. https://doi.org/10.1148/radiology.203.1.9122382

Kuhl, C. K., Gatsonis, C., Newstead, G., Snyder, B. S., An, N., Gareen, I. F., Bergin, J. T., Rahbar, H., Sung, J. S., Jacobs, J., Harvey, J. A., Nicholson, M. H., Ward, R. C., Holt, J., Prather, A., Miller, K., Schnall,

M. D., & Comstock, C. E. (2023). Incidence round screening performance among women with dense breasts undergoing abbreviated breast MRI and digital breast tomosynthesis (ECOG-ACRIN EA1141). *Journal of Clinical Oncology*, *41*(16_suppl), 10502–10502. https://doi.org/10.1200/JCO.2023.41.16_suppl.10502

Kuhl, C. K., Schrading, S., Leutner, C. C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., Kuhn, W., & Schild, H. H. (2005). Mammography, Breast Ultrasound, and Magnetic Resonance Imaging for Surveillance of Women at High Familial Risk for Breast Cancer. *Journal of Clinical Oncology*, *23*(33), 8469–8476. https://doi.org/10.1200/JCO.2004.00.4960

Kuhl, C., Weigel, S., Schrading, S., Arand, B., Bieling, H., König, R., Tombach, B., Leutner, C., Rieber-Brambs, A., Nordhoff, D., Heindel, W., Reiser, M., & Schild, H. H. (2010). Prospective Multicenter Cohort Study to Refine Management Recommendations for Women at Elevated Familial Risk of Breast Cancer: The EVA Trial. *Journal of Clinical Oncology*, *28*(9), 1450–1457. https://doi.org/10.1200/JCO.2009.23.0839

Kuo, Y.-T., Kazerouni, A. S., Park, V. Y., Surento, W., Hippe, D. S., Rahbar, H., & Partridge, S. C. (n.d.). *Validation of Fully-Automated Deep Learning-Based Fibroglandular Tissue Segmentation for Efficient and Reliable Quantitation of Background Parenchymal Enhancement in Breast MRI*.

Lam, D. L., Hippe, D. S., Kitsch, A. E., Partridge, S. C., & Rahbar, H. (2019). Assessment of Quantitative Magnetic Resonance Imaging Background Parenchymal Enhancement Parameters to Improve Determination of Individual Breast Cancer Risk. *Journal of Computer Assisted Tomography*, *43*(1), 85–92. https://doi.org/10.1097/RCT.0000000000000774

Lee, A. J., Cunningham, A. P., Tischkowitz, M., Simard, J., Pharoah, P. D., Easton, D. F., & Antoniou, A. C. (2016). Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genetics in Medicine*, *18*(12), 1190–1198. https://doi.org/10.1038/gim.2016.31

Lee, C. S., Monticciolo, D. L., & Moy, L. (2020). Screening Guidelines Update for Average-Risk and High-Risk Women. *American Journal of Roentgenology*, *214*(2), 316–323. https://doi.org/10.2214/AJR.19.22205

Lehman, C. D., Isaacs, C., Schnall, M. D., Pisano, E. D., Ascher, S. M., Weatherall, P. T., Bluemke, D. A., Bowen, D. J., Marcom, P. K., Armstrong, D. K., Domchek, S. M., Tomlinson, G., Skates, S. J., & Gatsonis, C. (2007). Cancer Yield of Mammography, MR, and US in High-Risk Women: Prospective Multi-Institution Breast Cancer Screening Study. *Radiology*, *244*(2), 381–388. https://doi.org/10.1148/radiol.2442060461

Lehman, C. D., & Smith, R. A. (2009). The Role of MRI in Breast Cancer Screening. *Journal of the National Comprehensive Cancer Network*, *7*(10), 1109–1115. https://doi.org/10.6004/jnccn.2009.0072

Lew, C. O., Harouni, M., Kirksey, E. R., Kang, E. J., Dong, H., Gu, H., Grimm, L. J., Walsh, R., Lowell, D. A., & Mazurowski, M. A. (2024). A publicly available deep learning model and dataset for segmentation of breast, fibroglandular tissue, and vessels in breast MRI. *Scientific Reports*, *14*(1), 5383. https://doi.org/10.1038/s41598-024-54048-2

Liao, G. J., Henze Bancroft, L. C., Strigel, R. M., Chitalia, R. D., Kontos, D., Moy, L., Partridge, S. C., & Rahbar, H. (2020). Background parenchymal enhancement on breast MRI: A comprehensive review. *Journal of Magnetic Resonance Imaging*, *51*(1), 43–61. https://doi.org/10.1002/jmri.26762

Lu, W. L., Jansen, L., Post, W. J., Bonnema, J., Van De Velde, J. C., & De Bock, G. H. (2009). Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: A meta-analysis. *Breast Cancer Research and Treatment*, *114*(3), 403–412. https://doi.org/10.1007/s10549-008-0023-4

Madu, C. O., Wang, S., Madu, C. O., & Lu, Y. (2020). Angiogenesis in Breast Cancer Progression, Diagnosis, and Treatment. *Journal of Cancer*, *11*(15), 4474–4494. https://doi.org/10.7150/jca.44313

Mazzola, E., Blackford, A., Parmigiani, G., & Biswas, S. (2015). Recent Enhancements to the Genetic Risk Prediction Model BRCAPRO. *Cancer Informatics*, *14s2*, CIN.S17292. https://doi.org/10.4137/CIN.S17292

McClintock, A. H., Golob, A. L., & Laya, M. B. (2020). Breast Cancer Risk Assessment. *Mayo Clinic Proceedings*, *95*(6), 1268–1275. https://doi.org/10.1016/j.mayocp.2020.04.017

Monticciolo, D. L., Newell, M. S., Hendrick, R. E., Helvie, M. A., Moy, L., Monsees, B., Kopans, D. B., Eby, P. R., & Sickles, E. A. (2017). Breast Cancer Screening for Average-Risk Women: Recommendations From the ACR Commission on Breast Imaging. *Journal of the American College of Radiology*, *14*(9), 1137–1143. https://doi.org/10.1016/j.jacr.2017.06.001

Moore, S. G., Shenoy, P. J., Fanucchi, L., Tumeh, J. W., & Flowers, C. R. (2009). Cost-effectiveness of MRI compared to mammography for breast cancer screening in a high risk population. *BMC Health Services Research*, *9*(1), 9. https://doi.org/10.1186/1472-6963-9-9

Müller-Schimpfle, M., Ohmenhaüser, K., Stoll, P., Dietz, K., & Claussen, C. D. (1997). Menstrual cycle and age: Influence on parenchymal contrast medium enhancement in MR imaging of the breast. *Radiology*, *203*(1), 145–149. https://doi.org/10.1148/radiology.203.1.9122383

Neeter, L. M. F. H., Robbe, M. M. Quirien., Van Nijnatten, T. J. A., Jochelson, M. S., Raat, H. P. J., Wildberger, J. E., Smidt, M. L., Nelemans, P. J., & Lobbes, M. B. I. (2023). Comparing the Diagnostic Performance of Contrast-Enhanced Mammography and Breast MRI: A Systematic Review and Meta-Analysis. *Journal of Cancer*, *14*(1), 174–182. https://doi.org/10.7150/jca.79747

Niell, B. L., Abdalah, M., Stringfield, O., Raghunand, N., Ataya, D., Gillies, R., & Balagurunathan, Y. (2021). Quantitative Measures of Background Parenchymal Enhancement Predict Breast Cancer Risk. *American Journal of Roentgenology*, *217*(1), 64–75. https://doi.org/10.2214/AJR.20.23804

Okuma, H., Masarwah, A., Istomin, A., Nykänen, A., Hakumäki, J., Vanninen, R., & Sudah, M. (2024). Increased background parenchymal enhancement on peri-menopausal breast magnetic resonance imaging. *European Journal of Radiology Open*, *13*, 100611. https://doi.org/10.1016/j.ejro.2024.100611

Ozanne, E. M., Drohan, B., Bosinoff, P., Semine, A., Jellinek, M., Cronin, C., Millham, F., Dowd, D., Rourke, T., Block, C., & Hughes, K. S. (2013). Which Risk Model to Use? Clinical Implications of the ACS MRI Screening Guidelines. *Cancer Epidemiology, Biomarkers & Prevention*, *22*(1), 146–149. https://doi.org/10.1158/1055-9965.EPI-12-0570

Partridge, S. C., Mullins, C. D., Kurland, B. F., Allain, M. D., DeMartini, W. B., Eby, P. R., & Lehman, C. D. (2010). Apparent Diffusion Coefficient Values for Discriminating Benign and Malignant Breast MRI Lesions: Effects of Lesion Type and Size. *American Journal of Roentgenology*, *194*(6), 1664–1673. https://doi.org/10.2214/AJR.09.3534

Partridge, S. C., Zhang, Z., Newitt, D. C., Gibbs, J. E., Chenevert, T. L., Rosen, M. A., Bolan, P. J., Marques, H. S., Romanoff, J., Cimino, L., Joe, B. N., Umphrey, H. R., Ojeda-Fournier, H., Dogan, B., Oh, K., Abe, H., Drukteinis, J. S., Esserman, L. J., Hylton, N. M., & For the ACRIN 6698 Trial Team and I-SPY 2 Trial Investigators. (2018). Diffusion-weighted MRI Findings Predict Pathologic Response in Neoadjuvant Treatment of Breast Cancer: The ACRIN 6698 Multicenter Trial. *Radiology*, *289*(3), 618–627. https://doi.org/10.1148/radiol.2018180273

Petticrew, M., Sowden, A., & Lister-Sharp, D. (2001). FALSE-NEGATIVE RESULTS IN SCREENING PROGRAMS: Medical, Psychological, and Other Implications. *International Journal of Technology Assessment in Health Care*, *17*(2), 164–170. https://doi.org/10.1017/S0266462300105021

Porter, G. J. R., Evans, A. J., Cornford, E. J., Burrell, H. C., James, J. J., Lee, A. H. S., & Chakrabarti, J. (2007). Influence of Mammographic Parenchymal Pattern in Screening-Detected and Interval Invasive Breast Cancers on Pathologic Features, Mammographic Features, and Patient Survival. *American Journal of Roentgenology*, *188*(3), 676–683. https://doi.org/10.2214/AJR.05.1950

Portnoi, T., Yala, A., Schuster, T., Barzilay, R., Dontchos, B., Lamb, L., & Lehman, C. (2019). Deep Learning Model to Assess Cancer Risk on the Basis of a Breast MR Image Alone. *American Journal of Roentgenology*, *213*(1), 227–233. https://doi.org/10.2214/AJR.18.20813

Pötsch, N., Vatteroni, G., Clauser, P., Helbich, T. H., & Baltzer, P. A. T. (2022). Contrast-enhanced Mammography versus Contrast-enhanced Breast MRI: A Systematic Review and Meta-Analysis. *Radiology*, *305*(1), 94–103. https://doi.org/10.1148/radiol.212530

Pujara, A. C., Mikheev, A., Rusinek, H., Gao, Y., Chhor, C., Pysarenko, K., Rallapalli, H., Walczyk, J., Moccaldi, M., Babb, J. S., & Melsaether, A. N. (2018). Comparison between qualitative and quantitative assessment of background parenchymal enhancement on breast MRI. *Journal of Magnetic Resonance Imaging*, *47*(6), 1685–1691. https://doi.org/10.1002/jmri.25895

Ready, K., Litton, J. K., & Arun, B. K. (2010). Clinical Application of Breast Cancer Risk Assessment Models. *Future Oncology*, *6*(3), 355–365. https://doi.org/10.2217/fon.10.5

Rosen, E. L., Smith-Foley, S. A., DeMartini, W. B., Eby, P. R., Peacock, S., & Lehman, C. D. (2007). BI-RADS MRI Enhancement Characteristics of Ductal Carcinoma In Situ: *MRI Characteristics of Ductal Carcinoma In Situ*. *The Breast Journal*, *13*(6), 545–550. https://doi.org/10.1111/j.1524-4741.2007.00513.x

Shattuck-Eidens, D., McClure, M., Rubano, T., Pruss, D., Staebell, M., Gumpper, K., Hulick, M., Holmen, J., Luce, M., Steele, L., Offit, K., Gilewski, T., Norton, L., Brown, K., Schulz, C., Giulotto, E., Zoli, W., Ravaioli, A., Nevanlinna, H., … Thomas, A. (n.d.). *BRCA1 Sequence Analysis in Women at High Risk for Susceptibility Mutations*.

Sickles, E. (1984). Mammographic features of “early” breast cancer. *American Journal of Roentgenology*, *143*(3), 461–464. https://doi.org/10.2214/ajr.143.3.461

Siegel, R. L., Kratzer, T. B., Giaquinto, A. N., Sung, H., & Jemal, A. (2025). Cancer statistics, 2025. *CA: A Cancer Journal for Clinicians*, *75*(1), 10–45. https://doi.org/10.3322/caac.21871

Smith, R. A., Andrews, K. S., Brooks, D., Fedewa, S. A., Manassaram-Baptiste, D., Saslow, D., & Wender, R. C. (2019). Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA: A Cancer Journal for Clinicians*, *69*(3), 184–210. https://doi.org/10.3322/caac.21557

Sogani, J., Mango, V. L., Keating, D., Sung, J. S., & Jochelson, M. S. (2021). Contrast-enhanced mammography: Past, present, and future. *Clinical Imaging*, *69*, 269–279. https://doi.org/10.1016/j.clinimag.2020.09.003

Stejskal, E. O., & Tanner, J. E. (1965). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, *42*(1), 288–292. https://doi.org/10.1063/1.1695690

Sung, J. S., Corben, A. D., Brooks, J. D., Edelweiss, M., Keating, D. M., Lin, C., Morris, E. A., Patel, P., Robson, M., Woods, M., Bernstein, J. L., & Pike, M. C. (2018). Histopathologic characteristics of background parenchymal enhancement (BPE) on breast MRI. *Breast Cancer Research and Treatment*, *172*(2), 487–496. https://doi.org/10.1007/s10549-018-4916-6

Surento, W., Fischer, R., Biswas, D., Hippe, D. S., Kazerouni, A. S., Kim, J. Y., Li, I., Gennari, J. H., Rahbar, H., & Partridge, S. C. (2025). Multiparametric MRI Markers Associated with Breast Cancer Risk in Women with Dense Breasts. *Cancers*, *17*(23), 3771. https://doi.org/10.3390/cancers17233771

Telegrafo, M., Rella, L., Stabile Ianora, A. A., Angelelli, G., & Moschetta, M. (2016). Breast MRI background parenchymal enhancement (BPE) correlates with the risk of breast cancer. *Magnetic Resonance Imaging*, *34*(2), 173–176. https://doi.org/10.1016/j.mri.2015.10.014

Tice, J. A., Cummings, S. R., Smith-Bindman, R., Ichikawa, L., Barlow, W. E., & Kerlikowske, K. (2008). Using Clinical Factors and Mammographic Breast Density to Estimate Breast Cancer Risk: Development and Validation of a New Predictive Model. *Annals of Internal Medicine*, *148*(5), 337–347. https://doi.org/10.7326/0003-4819-148-5-200803040-00004

Tyrer, J., Duffy, S. W., & Cuzick, J. (2004). A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in Medicine*, *23*(7), 1111–1130. https://doi.org/10.1002/sim.1668

Vartanian, R. K., & Weidner, N. (n.d.). *Correlation of Intratumoral Endothelial Cell Proliferation with Microvessel Density (Tumor Angiogenesis) and Tumor Cell Proliferation in Breast Carcinoma*.

Wang, H., Van Der Velden, B. H. M., Verburg, E., Bakker, M. F., Pijnappel, R. M., Veldhuis, W. B., Van Gils, C. H., & Gilhuijs, K. G. A. (2023). Assessing Quantitative Parenchymal Features at Baseline Dynamic Contrast-enhanced MRI and Cancer Occurrence in Women with Extremely Dense Breasts. *Radiology*, *308*(2), e222841. https://doi.org/10.1148/radiol.222841

Warner, E., Messersmith, H., Causer, P., Eisen, A., Shumak, R., & Plewes, D. (2008). Systematic Review: Using Magnetic Resonance Imaging to Screen Women at High Risk for Breast Cancer. *Annals of Internal Medicine*, *148*(9), 671–679. https://doi.org/10.7326/0003-4819-148-9-200805060-00007

Warner, E., Plewes, D. B., Hill, K. A., Causer, P. A., Zubovits, J. T., Jong, R. A., Cutrara, M. R., DeBoer, G., Yaffe, M. J., Messner, S. J., Meschino, W. S., Piron, C. A., & Narod, S. A. (n.d.). *Surveillance of BRCA1 and BRCA2 Mutation Carriers With Magnetic Resonance Imaging, Ultrasound, Mammography, and Clinical Breast Examination*.

Waters, E. A., Taber, J. M., McQueen, A., Housten, A. J., Studts, J. L., & Scherer, L. D. (2020). Translating Cancer Risk Prediction Models into Personalized Cancer Risk Assessment Tools: Stumbling Blocks and Strategies for Success. *Cancer Epidemiology, Biomarkers & Prevention*, *29*(12), 2389–2394. https://doi.org/10.1158/1055-9965.EPI-20-0861

Watt, G. P., Sung, J., Morris, E. A., Buys, S. S., Bradbury, A. R., Brooks, J. D., Conant, E. F., Weinstein, S. P., Kontos, D., Woods, M., Colonna, S. V., Liang, X., Stein, M. A., Pike, M. C., & Bernstein, J. L. (2020). Association of breast cancer with MRI background parenchymal enhancement: The IMAGINE case-control study. *Breast Cancer Research*, *22*(1), 138. https://doi.org/10.1186/s13058-020-01375-7

Weidner, N., Semple, J. P., Welch, W. R., & Folkman, J. (1991). Tumor Angiogenesis and Metastasis—Correlation in Invasive Breast Carcinoma. *New England Journal of Medicine*, *324*(1), 1–8. https://doi.org/10.1056/NEJM199101033240101

Yala, A., Mikhael, P. G., Strand, F., Lin, G., Smith, K., Wan, Y.-L., Lamb, L., Hughes, K., Lehman, C., & Barzilay, R. (2021). Toward robust mammography-based models for breast cancer risk. *Science Translational Medicine*, *13*(578), eaba4373. https://doi.org/10.1126/scitranslmed.aba4373

Zeppa, R. (1969). Vascular Response of the Breast to Estrogen[1]. *The Journal of Clinical Endocrinology & Metabolism*, *29*(5), 695–700. https://doi.org/10.1210/jcem-29-5-695