Risk Factors for Dense Breast in US Women: Insights from the Breast Cancer Surveillance Consortium

Abstract

Breast density is a key factor in breast cancer risk assessment, with denser tissue associated with increased likelihood of developing the disease. Previous research has identified age, body max index(BMI), and use of hormone replacement therapy (HRT) as the primary determinants of denser breasts. Using data from the Breast Cancer Surveillance Consortium (BCSC) of 63,577 mammograms in 2017, this study investigates the relationships between the odds of high breast density (BI-RADS ≥ 3) with these factors through multiple logistic regression models. Results showed that older age and BMI were significantly associated with lower odds of having dense breasts, while HRT use showed no significant independent effect. All pairwise interactions between these factors were statistically significant, suggesting complex relationships in their combined effects on breast density. However, the cross-sectional design, lack of recruitment details, and observational nature of the study limited its capacity for causal inference and broader generalization.

1 Introduction

Breast density is a well-established factor in assessing breast cancer risk, with denser breast tissue being associated with an increased likelihood of developing the disease (McCormack & dos Santos Silva, 2006; Bodewes et al., 2022). Despite advances in medical imaging techniques, such as mammography, breast density remains an important yet often underestimated determinant of breast cancer risk (Beidler et al., 2023). Understanding the factors that influence breast density is essential not only for improving early detection but also for addressing misconceptions that downplay its role as a key risk factor. Previous research has identified several factors, including age, body mass index (BMI), and the use of hormone replacement therapy (HRT), that may impact breast density, alongside more widely recognized factors such as family history. In particular, younger women, those with lower BMI, and those using HRT are more likely to exhibit higher breast density, as demonstrated by studies conducted in diverse cohorts from Norway, Denmark, China, and the US (Hou et al., 2013; Azam et al., 2018; Ji et al., 2021; Advani et al., 2021).

Based on these findings, we conducted a study to quantify the extent to which age group, BMI group, and HRT use contribute to the likelihood of having dense breasts in US women and explore whether these effects vary among women with different combinations of these risk factors, using publicly available data from the Breast Cancer Surveillance Consortium (BCSC). Our analysis centered around two questions: **(1) Is there an association between age group, BMI group, and HRT use with the likelihood of having higher breast density?** and **(2) If associations are present, do they vary among individuals with different combinations of these factors?**. Using multiple logistic regression models with participants in 2017, the most recent year in the database, we found that only age group and BMI group significantly impacted the likelihood of having denser breasts while the use of HRT was not, contrary to existing literature. Specifically, increasing age and BMI were associated with lower odds of having dense breasts. In addition, all pairwise interactions between predictor variables were significant, suggesting that the effects of these factors vary across individuals. However, the lack of information on recruitment methods, the reduction from a longitudinal to a single-year year study due to missing identifier variables, and the observational nature of the study limited its capacity for causal inference and broader generalization.

2 Materials & Methods

2.1 Dataset

Data for this study were obtained from the BCSC, a collaborative network that includes six active breast imaging registries, three historic registries, and a research-focused enterprise, accessible at `http://www.bcsc-research.org/`. The BCSC aims to evaluate and improve the quality as well as the delivery of breast cancer screening and associated outcomes across the US. Each registry performs annual linkages with tumor and pathology registries in its geographic region, supported by a central Statistical Coordinating Center. This study used the dataset from the risk factor section of the BCSC database updated in March 2020. The dataset includes 6,788,436 mammograms recorded between January 2005 and December 2017, along with participant characteristics associated with breast cancer risk, such as age, race/ethnicity, family history of breast cancer, age at menarche, age at first birth, breast density, use of hormone replacement therapy, menopausal status, body mass index, history of biopsy, and prior breast cancer diagnosis. To create this dataset, one mammogram per woman was selected for each calendar year and year of age. When both screening and diagnostic mammograms were available for the same woman in a given year, screening mammograms were preferentially included.

2.2 Data Processing & Analysis

This study investigates the relationship between the likelihood of having dense breast and three key risk factors, including age, BMI, and the use HRT so any variables other than *age_group_5_years* (i.e., the age of the participant at 5-year intervals, categorized into 13 levels from 18 to over 85 years),

bmi_group (i.e., body mass index in kg/m^2, categorized into 4 levels from 10 to over 35), *current_hrt* (i.e., current use of HRT, categorized into 2 levels of yes and no) and *BIRADS_breast_density* (i.e., breast density levels, categorized into 4 levels according to the Breast Imaging Reporting and Data System) were excluded. A new binary variable, *breast_density_group*, was created to classify breast density into two broader categories: *Dense* (BI-RADS categories 3 and 4, coded as 1) and *Not Dense* (BI-RADS categories 1 and 2, coded as 0) (Table A1).

The dataset was constructed by selecting one mammogram per woman for each calendar year and year of age, which introduced the possibility of repeated measures for some participants and therefore, violated the assumption of independence. To address this issue, only observations from the most recent year, 2017, were retained. In addition, observations with missing or unknown values (coded as 9) in any of the variables, except for *age_group_5_years*, were also excluded, resulting in a final dataset of 5 variables with 63,577 complete observations.

Multiple logistic regression models were used to analyze the binary response variable *breast_density_group* with categorical predictor variables *age_group_5_years*, *bmi_group*, and *current_hrt* (Model 1). Pairwise interactions among the predictors were explored to evaluate potential combined effects (Model 2). Since all predictors were categorical, the assumption of linearity was inherently satisfied, eliminating the need for empirical logit plots. The significance of individual predictors and interactions was assessed using the Likelihood Ratio Test (LRT), while the overall robustness was evaluated using the Hosmer-Lemeshow test. Statistical significance was determined at a p-value threshold of < 0.05.

3 Results

3.1 Descriptive Analysis

Descriptive analysis reveals the distribution of participant characteristics in 2017 (Tables B1 & B2). Most participants were middle-aged or older, with the 50–59 age group representing 33.36%, while younger participants (18–39) and those aged 75 and above made up 5.30% and 11.79%, respectively. BMI was most concentrated in the 10–24.99 (30.74%) and 25–29.99 (28.91%) categories, with fewer participants in the BMI over 35 group (18.78%). HRT usage was rare, at 7.38%, and 53.09% of participants had not-dense breasts (BI-RADS ≤ 2), while 46.91% had dense breasts (BI-RADS ≥ 3). Breast density was more prevalent in those under 50, and not-dense breasts were more common in those over 55, with lower BMI associated with higher breast density (Figure B1). Contrary to expectations, HRT use did not appear to increase breast density (Figure B1). BMI showed no consistent pattern with age (Figure B2). The descriptive analysis suggests that there are no apparent confounding effects between age, BMI, and HRT use on breast density.

3.2 Multiple Logistic Models

The logistic regression analysis identified significant risk factors for breast density and highlighted the role of interaction effects among these factors (Table C1). In the model without interactions (Model 1), age, BMI, and current HRT use were examined. Age demonstrated a strong negative relationship with breast density, with individuals in older age groups having significantly lower odds of higher breasts density (OR = 0.863, $p < 0.001$). Similarly, BMI was also negatively associated with breast density, with individuals in higher BMI groups showing significantly lower odds of having dense breasts (OR = 0.596, $p < 0.001$). Although the effect of current HRT use on breast density was not statistically significant (OR = 0.972, $p = 0.375$), it also demonstrated a negative association, suggesting that HRT use may be linked to lower odds of having dense breasts.

Model 2, which incorporated pairwise interaction effects, provided additional insights and slightly improved model fit, as indicated by a minor reduction in AIC (81580 vs. 81670). Age (OR = 0.809, $p < 0.001$) and BMI (OR = 0.493, $p < 0.001$) remained significant predictors, with effects consistent with those observed in Model 1. All pairwise interactions were statistically significant.

The interaction plot between age and BMI groups revealed a downward trend in the likelihood of having dense breasts across all BMI categories with increasing age, reinforcing the negative association of both factors with breast density (Figure C1). The slopes were steeper for lower BMI groups, indicating that individuals with lower BMI had higher breast density at younger ages, but this likelihood decreased significantly with age. Although individuals in higher BMI groups were less likely to have dense breasts at younger ages, this likelihood remained relatively higher as they aged. The non-parallel lines suggest an interaction between age and BMI, supported by the statistically significant interaction term (OR = 1.029, $p < 0.001$) (Table C1).

The interaction plot between age and HRT use showed a similar overall downward trend (Figure C2). Non-HRT users were more likely to develop dense breasts at younger ages, but this likelihood decreased more rapidly with age. In contrast, HRT users were less likely to have dense breasts at younger ages, and although the likelihood still decreased with age, it remained higher than that of non-users over the age of 75. The non-parallel lines further suggest an interaction between age and HRT use, supported by the statistically significant interaction term (OR = 1.055, $p < 0.001$) (Table C1).

The interaction plot between BMI and HRT use also showed a downward trend (Figure C3). At lower BMI levels, HRT users and non-users had similar high likelihoods of having dense breasts. However, as BMI increased, a divergence emerged, with non-users maintaining a relatively higher likelihood of dense breasts compared to HRT users. The non-parallel lines suggest an interaction between BMI and HRT use, supported by the statistically significant interaction term (OR = 0.898, $p = 0.001$) (Table C1).

These findings were further validated by LRT analyses, where age ($p < 0.001$) and BMI ($p < 0.001$) remained significant predictors in both models, with and without interaction terms. HRT use was not significant in either model ($p = 0.375$ in Model 1 and $p = 0.102$ in Model 2). The interaction terms were statistically significant ($p < 0.001$). However, the Hosmer-Lemeshow test indicated a lack of fit in both models, with p-values < 0.001.

4 Discussions

These findings align with prior researcch, highlighting the multi-factorial nature of breast density. Younger age and lower BMI are significant predictors of higher breast density, consistent with established evidence. Younger women often have dense fibroglandular tissue, which diminishes with age due to breast involution (Maskarinec et al., 2016). Similarly, higher BMI was associated with reduced breast density, potentially due to the increased proportion of adipose tissue. In contrast, HRT usage demonstrated negative association with breast density, suggesting a lower likelihood of having dense breast when used HRT, which is contrast to the literature (Hou et al, 2013; Azam et al., 2018). Interaction terms indicate the combined effects of these factors and the nuanced influence of HRT use on other factors, regardless of its standalone insignificant. The protective effect of higher BMI against dense tissue weakens with age, while metabolic or hormonal differences may explain attenuated HRT effects in low BMI groups. These insights underscore the importance of incorporating these complex interactions into breast cancer risk models, potentially improving personalized screening and risk stratification strategies.

One limitation of this study is the absence of detailed information on recruitment methods, which restricts understanding of sample representativeness and potential biases. The reduction to a single-year study due to missing identifier variables further weakened the analysis, especially in assessing temporal changes. This is particularly relevant for variables like HRT usage, as hormone therapy often spans 2–5 years, with initial three-month trials to determine suitability. Without longitudinal data, the full impact of HRT on breast density may not be accurately reflected, potentially underestimating or misrepresenting its association. Another limitation pertains to the causal inference of the study results. Since this study is observational without any treatment assignment, the results can only reflect the relationships between these three risk factors with the likelihood of having higher breast density.

Future research should request the identifier variables of the participants and information regarding recruitment methods to improve the analysis and the extend the impact of the conclusions. In addition, other factors such as genetic predispositions and lifestyle influences, to develop a more comprehensive understanding of breast density determinants.

5 Appendix

5.1 Appendix A - Categories of Variables

Variable	Description	Level
age_group_5_years	Age (years) in 5 year groups	1 = Age 18-29 2 = Age 30-34 3 = Age 35-39 4 = Age 40-44 5 = Age 45-49 6 = Age 50-54 7 = Age 55-59 8 = Age 60-64 9 = Age 65-69 10 = Age 70-74 11 = Age 75-79 12 = Age 80-84 13 = Age $\geq$ 85
current_hrt	Use of hormone replacement therapy	0 = No 1 = Yes 9 = Unknown
bmi_group	Body mass index (kg/m2)	1 = 10-24.99 2 = 25-29.99 3 = 30-34.99 4 = 35 or more 9 = Unknown
BIRADS_breast_density	BI-RADS breast density	1 = Almost entirely fat 2 = Scattered fibroglandular densities 3 = Heterogeneously dense 4 = Extremely dense 9 = Unknown or different measurement system
breast_density_group	Classification of breast density	0 = Not Dense 1 = Dense

Table A1. Description of the levels of each variables

5.2 Appendix B - Descriptive Analysis

Age Group	Count	BMI Group	Count
Age 18-29	220 (0.35%)	10-24.99	19544 (30.74%)
Age 30-34	1057 (1.66%)	25-29.99	18378 (28.91%)
Age 35-39	2092 (3.29%)	30-34.99	13715 (21.57%)
Age 40-44	5783 (9.10%)	Over 35	11940 (18.78%)
Age 45-49	8107 (12.75%)		
Age 50-54	10652 (16.75%)		
Age 55-59	10563 (16.61%)		
Age 60-64	6333 (9.96%)		
Age 65-69	6139 (9.66%)		
Age 70-74	5134 (8.08%)		
Age 75-79	3688 (5.80%)		
Age 80-84	2348 (3.69%)		
Age 85+	1461 (2.30%)		

Table B1. Distribution of Age Group and BMI Group with counts and proportions

HRT Status	Count	Breast Density Group	Count
No	58883 (92.62%)	Not Dense	33755 (53.09%)
Yes	4694 (7.38%)	Dense	29822 (46.91%)

Table B2. Distribution of HRT Status and Breast Density Group with counts and proportions

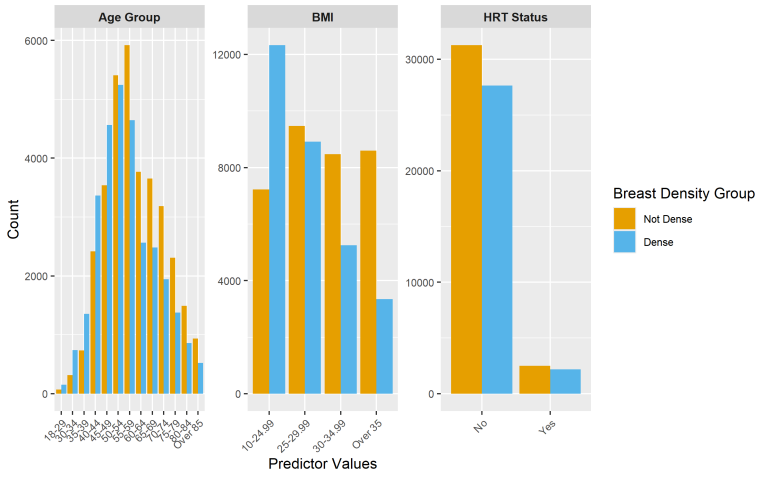

Figure B1. Distribution of participants by Breast Density, Age, BMI, and HRT groups

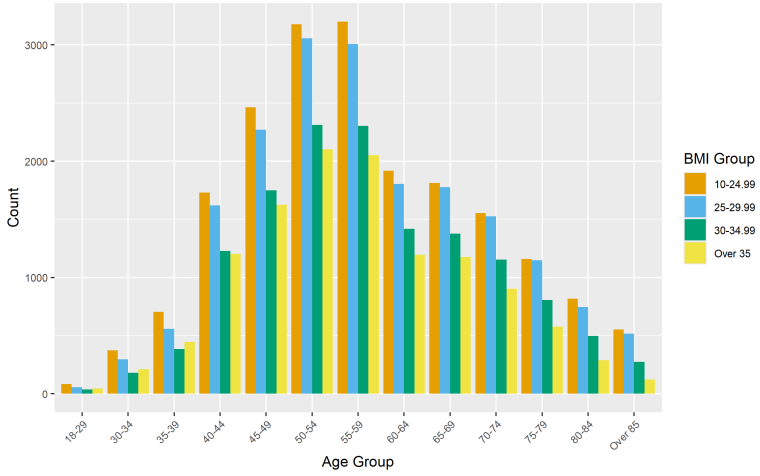

Figure B2. Distribution of participants by Age and BMI groups

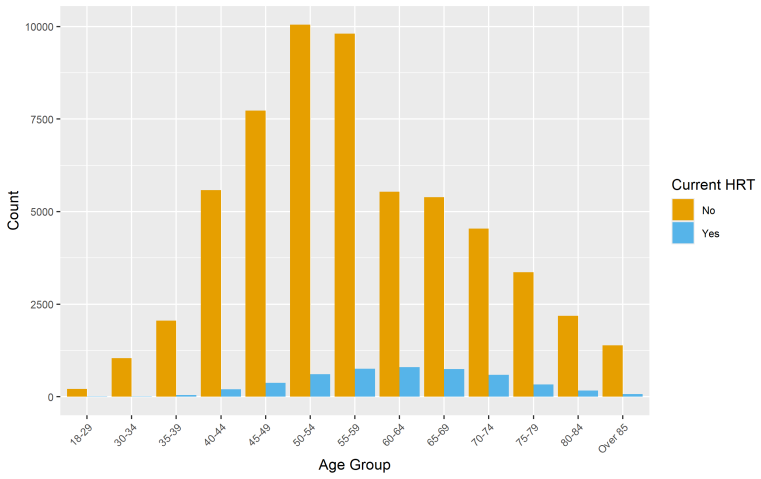

Figure B3. Distribution of participants by Age and HRT groups

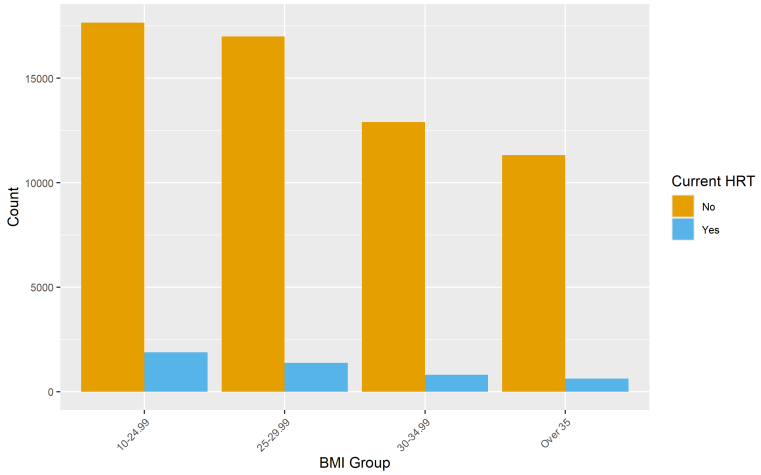

Figure B4. Distribution of participants by BMI and HRT groups

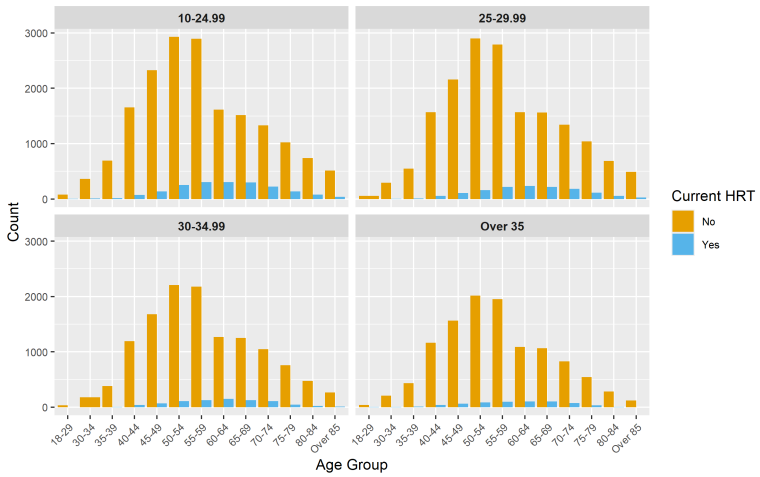

Figure B5. Distribution of participants by Age, BMI, and HRT groups

5.3 Appendix C - Multiple Logistic Models Results

	Model 1 (No Interactions)				Model 2 (With Interactions)			
Variable	**Coeff.**	**Odds Ratio**	**P-val (Wald)**	**P-val (LRT)**	**Coeff.**	**Odds Ratio**	**P-val (Wald)**	**P-val (LRT)**
Age	-0.147	0.863	< 0.001	< 0.001	-0.212	0.809	< 0.001	
HRT	-0.028	0.972	0.375	0.375	-0.232	0.793	0.102	
BMI	-0.518	0.596	< 0.001	< 0.001	-0.708	0.493	< 0.001	
Age $\times$ HRT					0.053	1.055	< 0.001	< 0.001
Age $\times$ BMI					0.028	1.029	< 0.001	< 0.001
BMI $\times$ HRT					-0.107	0.898	0.001	< 0.001

Table C1. Summary of Models

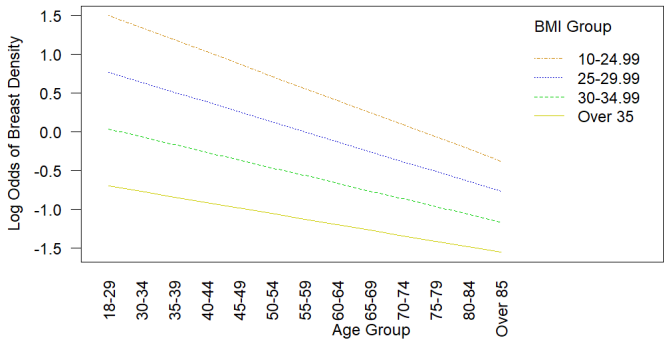

Figure C1. Interaction between Age and BMI groups

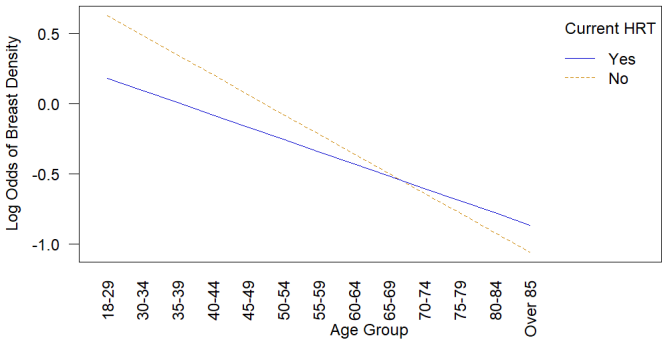

Figure C2. Interaction between Age and HRT groups

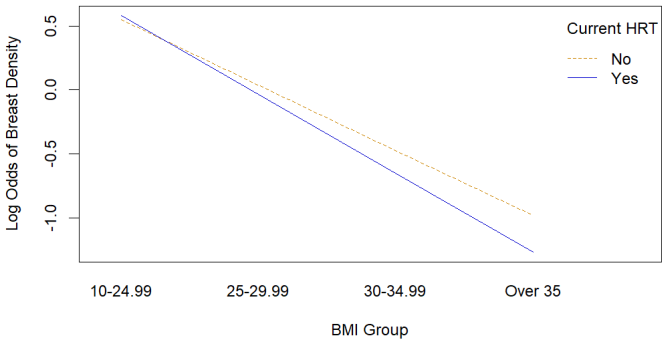

Figure C3. Interaction between BMI and HRT groups

6 Reference

1. Advani, S. M., Zhu, W., Demb, J., et al. (2021). Association of breast density with breast cancer risk among women aged 65 years or older by age group and body mass index. *JAMA Network Open, 4(8)*. https://doi.org/10.1001/jamanetworkopen.2021.22810.

2. Azam, S., Lange, T., Huynh, S., et al. (2018). Hormone replacement therapy, mammographic density, and Breast Cancer Risk: A cohort study. *Cancer Causes & Control, 29(6)*, 495–505. https://doi.org/10.1007/s10552-018-1033-0.

3. Beidler, L. B., Kressin, N. R., Wormwood, J. B., Battaglia, T. A., Slanetz, P. J., & Gunn, C. M. (2023). Perceptions of breast cancer risks among women receiving Mammograph screening. *JAMA Network Open, 6(1)*. https://doi.org/10.1001/jamanetworkopen.2022.52209

4. Bodewes, F. T. H., van Asselt, A. A., Dorrius, M. D., Greuter, M. J. W., & de Bock, G. H. (2022). Mammographic breast density and the risk of breast cancer: A systematic review and meta-analysis. *The Breast, 66*, 62–68. https://doi.org/10.1016/j.breast.2022.09.007

5. Hou, N., Hong, S., Wang, W., Olopade, O. I., Dignam, J. J., & Huo, D. (2013). Hormone replacement therapy and breast cancer: Heterogeneous risks by race, weight, and breast density. *Journal of the National Cancer Institute, 105(18)*, 1365–1372. https://doi.org/10.1093/jnci/djt207.

6. Ji, Y., Li, B., Zhao, R., Zhang, Y., Liu, J., & Lu, H. (2021). The relationship between breast density, age, and mammographic lesion type among Chinese breast cancer patients from a large clinical dataset. *BMC Medical Imaging, 21(1)*. https://doi.org/10.1186/s12880-021-00565-9.

7. Maskarinec, G., Ju, D., Horio, D., Loo, L. W., & Hernandez, B. Y. (2016). Involution of breast tissue and mammographic density. *Breast Cancer Research, 18(1)*. https://doi.org/10.1186/s13058-016-0792-3.

8. McCormack, V. A., & dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention, 15(6)*, 1159–1169. https://doi.org/10.1158/1055-9965.epi-06-0034.

9. The Breast Cancer Surveillance Consortium and its data collection and sharing activities are funded by the National Cancer Institute (P01CA154292). Downloaded 12/01/2025 from the Breast Cancer Surveillance Consortium Website - http://www.bcsc-research.org/.