Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Exploratory Study of Breast Cancer Risk Stratification Based on the Interaction Between First- Degree Family History and BMI

Yiwen Cao
Fudan University
Kefang Wang
Shanghai Health and Medical Center
Liqun Shi
Shanghai Health and Medical Center
Tiantian Zhang
Fudan University
Rui Ding
Fudan University
Xiaohui Zhou
doczxh@126.com

Shanghai Health and Medical Center

Research Article

Keywords: Breast cancer, Family history, Body mass index, Statistical interaction, GAIL model, Exploratory risk stratification

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11074330/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background Breast cancer remains the most prevalent malignant tumor among women worldwide, imposing a heavy public health burden in China. Precision risk stratification is the cornerstone of early breast cancer prevention. However, most existing risk prediction models are developed based on Western populations and frequently overlook the interaction between first-degree family history and body mass index (BMI), limiting their applicability in Chinese women.

Methods This cross-sectional study included 108 women who underwent physical examination and breast cancer risk assessment at a tertiary hospital in Wuxi, China, from 2021 to 2024. The outcome was the GAIL model-estimated 5-year breast cancer probability, analyzed as a continuous value and as low-, intermediate-, or high-risk categories; it was not an observed cancer event. Ordinal logistic regression evaluated associations of BMI and first-degree family history with higher GAIL-derived risk categories. Interaction was examined using age-adjusted stratified analyses, the Breslow-Day test, and an ordinal logistic model containing a BMI × family history term. The three-tier framework was assessed descriptively in the development sample; no time-to-event analysis was performed.

Results Overweight/obesity showed a modest association with a higher GAIL-derived category (ordinal OR = 1.20, 95% CI: 0.99–1.45), and positive first-degree family history was associated with a higher category (ordinal OR = 1.22, 95% CI: 1.01–1.47). The interaction term was positive (OR = 2.87, 95% CI: 1.24–6.64, P = 0.014); among women with a positive family history, the ordinal OR for overweight/obesity was 3.21 (95% CI: 1.76–5.85). In the same sample used to construct the framework, 70.00% of the proposed high tier, 42.86% of the intermediate tier, and 0% of the low tier were classified as high risk by the GAIL model. These proportions describe apparent separation of a model-derived outcome and should not be interpreted as observed 5-year incidence, relative risk, or a hazard ratio.

Conclusions First-degree family history and BMI may interact in relation to GAIL-estimated risk classification. The proposed framework is hypothesis-generating rather than a validated clinical prediction tool. Bootstrap internal validation using participant-level data and prospective external validation against observed breast cancer outcomes are required before clinical application.

Introduction

Breast cancer (BC) remains the most prevalent malignant tumor among women worldwide, imposing a substantial socioeconomic and public health burden. According to the *2020 Global Cancer Statistics*, there are approximately 2.3 million BC cases diagnosed globally each year, with China accounting for 420,000 cases, nearly one-fifth of the global incidence[1]. Beyond its high incidence, BC exhibits remarkable preventability: emerging evidence indicates that early prevention strategies can reduce BC incidence by 30%–50%, underscoring the urgency of identifying high-risk populations to enable precise interventions [2,3]. This need is further reinforced by policy imperatives: *China's Healthy China Action—Cancer Prevention and Control Implementation Plan (2023–2030)*[4] explicitly designates promoting cancer

screening and early diagnosis and treatment as a key priority, highlighting the policy-driven demand for BC risk population stratification tailored to the Chinese context.

Nevertheless, current BC risk assessment and prevention strategies, particularly those for Chinese populations, suffer from three critical limitations. First, mainstream risk prediction models (e.g., the Gail model, Tyrer-Cuzick model, and BCSC model) are primarily developed based on Western cohorts, leading to sub-optimal performance in Chinese women[5,6].For instance, the BCSC model significantly underestimates the invasive breast cancer risk for young women (< 45 years old) and women with non-dense breasts, which is particularly problematic in China where the age at peak incidence is 5–10 years earlier than in Western countries and the proportion of young patients is higher[5]. Furthermore, the actual risk of Chinese women with non-dense breasts differs from that of Western population[s6]. Similarly, despite incorporating comprehensive risk factors, the Tyrer-Cuzick model requires complex inputs such as third-degree relative family history and genetic test results, which are incompatible with China's primary healthcare environment; moreover, the model excludes critical risk factors for the Chinese population, such as environmental exposures, further limiting screening precision[7,8].

Second, a one-size-fits-all screening strategy exacerbates the problem of misallocated resources: high-risk populations often lack timely surveillance, while low-risk individuals may receive unnecessary screening, wasting limited healthcare resources. Third, and most notably, existing research and clinical guidelines, such as the National Comprehensive Cancer Network guidelines, treat family history and body mass index (BMI) as independent risk factors, overlooking their potential interactions[9,10]. This limitation restricts the precision of risk stratification. As a modifiable risk factor, BMI has been demonstrated to amplify breast cancer risk in individuals with genetic susceptibility by promoting estrogen synthesis and inhibiting homologous recombination repair (HRR) pathways[11]. However, whether this synergistic effect applies to individuals with a first-degree family history of breast cancer (FHF) remains unexplored in the Chinese population[12–14].

To address this gap, we examined whether BMI and first-degree family history were jointly associated with the GAIL model-estimated 5-year breast cancer risk among Chinese women. The objectives were to: (1) quantify associations with the GAIL-derived ordinal risk category; (2) test a BMI × family history interaction after adjustment for age; and (3) develop a parsimonious, exploratory stratification framework based on routinely available variables. Because the outcome was a predicted probability rather than an observed cancer event, all effect estimates were interpreted as associations with model-estimated risk classification, not as etiologic or prognostic effects on breast cancer incidence.

Materials and Methods

Study Design, Participants and Data Collection

This cross-sectional study was conducted at the physical examination center of a tertiary hospital in Wuxi, Yangtze River Delta region of China, between 2021 and 2024. Convenience sampling was used to

recruit female participants. Eligible participants were women aged ≥ 18 years who completed comprehensive physical examinations (including anthropometric measurements and breast-related examinations) and a structured breast cancer genetic testing questionnaire, and provided written informed consent. Individuals were excluded if they had a prior history of malignant tumors (e.g., breast, ovarian cancer), missing key data (BMI, family history, breast cancer risk score), or severe cardiac, hepatic, renal dysfunction or psychiatric disorders that prevented study participation. A total of 108 eligible participants were included in the final analysis (participant flow diagram in Supplementary Fig. 1).

Data were obtained from the center's standardized Electronic Health Records (EHRs), physical examination records, and a structured breast cancer genetic testing questionnaire. The original Chinese-language questionnaire was developed by the research team specifically for this study and has not previously been published. A faithful English-language translation is provided as Supplementary File 1. The questionnaire collected reproductive and hormonal history (including age at menarche and first childbirth, pregnancy and delivery history, breastfeeding, oral contraceptive use, menopausal status, and hormone replacement therapy); breast examination and benign breast disease history (breast ultrasonography, mammography, and history of benign breast surgery or biopsy); detailed family history of breast cancer and selected related cancers among first-, second-, and third-degree relatives, including age at diagnosis and genetic testing results in affected relatives; and personal breast cancer history, treatment, and genetic testing information. The questionnaire was designed for broader breast cancer genetic risk assessment, and not all questionnaire fields were used in the present analysis. BMI was calculated from height and weight measured during the physical examination and recorded in the EHRs rather than from the questionnaire. Uniformly trained researchers extracted objective clinical data from the EHRs, and all data were linked using anonymized identification numbers.

Variable Definition

The primary outcome was the GAIL model-estimated 5-year probability of breast cancer, expressed continuously and categorized as low (< 0.8%), intermediate (0.8%–1.19%), or high (≥ 1.2%) risk. The GAIL estimate is a model-derived surrogate rather than an observed clinical endpoint. It incorporates age, age at menarche, age at first childbirth, first-degree family history, breast biopsy history, and atypical hyperplasia history. Accordingly, analyses involving first-degree family history may partly reflect its direct inclusion in the outcome-generating model and were interpreted cautiously.

Core explanatory variables were BMI category and first-degree family history of breast cancer. BMI was calculated as weight (kg)/height (m^2) and classified per the Guidelines for the Prevention and Control of Overweight and Obesity in Chinese Adults: normal (18.5 ≤ BMI < 24.0, coded 0), overweight (24.0 ≤ BMI < 28.0, coded 1), and obese (BMI ≥ 28.0, coded 2). A binary classification (normal = 0, overweight/obese = 1) was used for sensitivity analysis. First-degree family history was categorized as: no cancer history (coded 0), breast cancer-only history (coded 1), and mixed cancer history (concurrent breast and other malignant tumors, coded 2), with a binary classification (negative = 0, positive = 1) for sensitivity

analysis. This three-category static definition aligns with international standards and captures the gradient of genetic susceptibility associated with different family history profiles, balancing clinical operability and risk discrimination[15]. The risk-score and GAIL-based classification approach was informed by prior Chinese high-risk screening work, regional risk-assessment research, and the NCI Breast Cancer Risk Assessment Tool[16,21,22].

Data Preprocessing and Statistical Analysis

Core analytical variables (age, BMI category, first-degree family history, continuous breast cancer risk, risk category) were extracted from the raw dataset, with irrelevant fields (e.g., physical examination ID) excluded. Column names were standardized, and missing values/outliers (BMI category 0–2, risk value 0–5%) were checked; no missing values were found, and outliers were excluded. Numeric codes of categorical variables were converted to labels for interpretation.

All analyses were conducted in Python 3.10 and R 4.4.0 using two-sided tests with $P < 0.05$. Continuous GAIL estimates were summarized by mean, standard deviation, median, and interquartile range, and compared across categories using the Kruskal-Wallis test. Cohen's Kappa assessed agreement between the original and adjusted classifications; it was not interpreted as predictive stability. Fisher's exact test and proportional-odds ordinal logistic regression were used to examine associations with higher GAIL-derived risk categories, with results reported as ordinal odds ratios (ORs) and 95% confidence intervals (CIs). Interaction was evaluated through cross-tabulation, age-adjusted Mantel-Haenszel analysis, the Breslow-Day test, and an age-adjusted ordinal logistic model containing BMI, family history, and their interaction term. The proposed three-tier framework was evaluated descriptively by the proportion classified as GAIL high risk in each tier. Kaplan-Meier and Cox models were not used because the study contained neither follow-up time nor observed cancer events.

Results

Baseline Characteristics of Study Participants

A total of 108 female participants were enrolled in this study, with a median age of 40 years (interquartile range: 35–50 years). The baseline characteristics are presented in Table 1. Regarding BMI categories, 37 participants (34.26%) had normal BMI, 42 (38.89%) were overweight, and 29 (26.85%) were obese, with 71 participants (65.74%) being overweight or obese. For first-degree family history, 26 participants (24.07%) had no family history of cancer, 45 (41.67%) had breast cancer-only family history, and 37 (34.26%) had mixed cancer family history, with 82 participants (75.93%) having a positive family history. Based on the GAIL model, 35 participants (32.41%) were classified as low-risk, 28 (25.93%) as intermediate-risk, and 45 (41.67%) as high-risk.

Table 1
Baseline characteristics

Characteristics	No. of participants (%)			OR (95% CI)
	High-risk group (n = 44)	Low-risk group (n = 64)	Total (N = 108)	
Age, y				
<35	4 (9.1)	18 (28.1)	22 (20.4)	1 [Reference]
35–44	27 (61.4)	24 (37.5)	51 (47.2)	5.06 (1.59–16.13)
45–54	10 (22.7)	12 (18.8)	22 (20.4)	3.75 (1.01–13.94)
≥ 55	3 (6.8)	10 (15.6)	13 (12.0)	1.35 (0.28–6.52)
BMI				
18.5–23.9 (Normal, coded 0)	25 (56.8)	45 (70.3)	70 (64.8)	1 [Reference]
24.0-27.9 (Overweight, coded 1)	11 (25.0)	11 (17.2)	22 (20.4)	1.80 (0.70–4.63)
≥ 28.0 (Obese, coded 2)	8 (18.2)	8 (12.5)	16 (14.8)	1.80 (0.61–5.31)
First-degree family history of breast cancer				
No (coded 0)	3 (6.8)	33 (51.6)	36 (33.3)	1 [Reference]
Breast cancer-only (coded 1)	12 (27.3)	16 (25.0)	28 (25.9)	8.25 (2.09–32.57)
Mixed cancer (coded 2)	29 (65.9)	15 (23.4)	44 (40.7)	21.27 (5.89–76.71)

Distribution of GAIL-Estimated Risk and Classification Agreement

The distribution of GAIL-estimated 5-year probabilities across the three predefined categories is shown in Table 2. Mean estimates were 0.0055 ± 0.0016 in the low category, 0.0103 ± 0.0011 in the intermediate category, and 0.0152 ± 0.0020 in the high category (Kruskal-Wallis $\chi^2 = 87.23$, df = 2, $P < 0.001$). This separation is expected because the categories were defined directly from the same continuous GAIL estimate and does not constitute independent validation of the model.

Table 2
Distribution of GAIL-estimated 5-year breast cancer probabilities across predefined risk categories

GAIL-Derived Risk Category	Sample Size (n, %)	GAIL 5-Year Probability (Mean ± SD)	GAIL 5-Year Probability (Median, IQR)
Low-risk	35, 32.41	0.0055 ± 0.0016	0.0051, (0.0042, 0.0068)
Intermediate-risk	28, 25.93	0.0103 ± 0.0011	0.0102, (0.0095, 0.0110)
High-risk	45, 41.67	0.0152 ± 0.0020	0.0150, (0.0138, 0.0165)
Total	108, 100.00	0.0108 ± 0.0041	0.0100, (0.0065, 0.0152)
Group Comparison (Kruskal-Wallis H Test)	$\chi^2 = 87.23$, df = 2, $P < 0.001$		
Agreement Between Original and Adjusted Classifications (Cohen's Kappa)	Kappa = 0.83 (95% CI: 0.74–0.91), $P < 0.001$		

Agreement between the original and adjusted classifications was high (Cohen's Kappa = 0.83, 95% CI: 0.74–0.91, $P < 0.001$). Kappa quantifies classification agreement only; it does not establish calibration, discrimination for observed breast cancer, reproducibility in new samples, or overall model stability.

Association Analysis

Fisher's exact tests indicated that BMI category and first-degree family history were associated with the three-level GAIL-derived classification (both reported $P < 0.001$). In ordinal logistic regression (Table 3), overweight/obesity showed a modest association with a higher category (OR = 1.20, 95% CI: 0.99–1.45), whereas positive family history was associated with a higher category (OR = 1.22, 95% CI: 1.01–1.47). Obesity and mixed family history had the largest subgroup estimates (OR = 1.48, 95% CI: 1.08–2.03; and OR = 1.56, 95% CI: 1.12–2.17, respectively). These estimates refer to the odds of occupying a higher GAIL-derived category, not to breast cancer incidence.

Table 3
Univariable ordinal logistic regression of BMI and first-degree family history with higher GAIL-derived risk category

Predictor	Category/Subgroup	Reference Group	Sample Size (n)	Ordinal OR (95% CI)	P-Value
BMI Classification	Normal (18.5–23.9)	—	37	1.00 (Reference)	—
	Overweight (24.0–27.9)	Normal (18.5–23.9)	42	1.32 (0.98–1.78)	0.068
	Obese (≥ 28.0)	Normal (18.5–23.9)	29	1.48 (1.08–2.03)	0.015
	Overweight/Obese (≥ 24.0)	Normal (18.5–23.9)	71	1.20 (0.99–1.45)	< 0.001
First-Degree Family History of Breast Cancer	Negative (no malignant tumor history)	—	26	1.00 (Reference)	—
	Positive – Breast cancer-only	Negative	45	1.38 (1.01–1.89)	0.042
	Positive – Mixed cancer (breast + others)	Negative	37	1.56 (1.12–2.17)	0.008
	Positive (all subtypes)	Negative	82	1.22 (1.01–1.47)	< 0.001

Figure 1 shows a rightward shift in GAIL-estimated probabilities among participants with overweight/obesity or positive family history. The proportion classified as GAIL high risk was 56.34% in the overweight/obese group versus 21.62% in the normal-BMI group, and 52.44% in the family-history-positive group versus 7.69% in the negative group. These are descriptive differences in a model-derived classification.

Interaction Analysis Between BMI and First-Degree Family History

Table 4 summarizes the distribution of GAIL-derived categories across the four BMI × family history groups (Pearson $\chi^2 = 67.76$, df = 6, $P < 0.001$). All participants with normal BMI and negative family history were in the low category, whereas 70.00% of those with overweight/obesity and positive family history were in the high category. The pattern was compatible with effect modification, although it cannot be interpreted as a difference in observed cancer incidence.

Table 4
Distribution of GAIL-derived risk categories across BMI × first-degree family history subgroups

BMI Category × First-Degree Family History	GAIL-Derived Risk Category			Total (n, %)
	Low-risk (n, %)	Intermediate-risk (n, %)	High-risk (n, %)	
Normal BMI (18.5–23.9) + Family history negative	8, 100.00	0, 0.00	0, 0.00	8, 7.41
Normal BMI (18.5–23.9) + Family history positive	13, 44.83	8, 27.59	8, 27.59	29, 26.85
Overweight/Obese (≥ 24.0) + Family history negative	0, 0.00	16, 76.19	5, 23.81	21, 19.44
Overweight/Obese (≥ 24.0) + Family history positive	12, 24.00	3, 6.00	35, 70.00	50, 46.30
Total	**33, 30.56**	**27, 25.00**	**48, 44.44**	**108, 100.00**

After age adjustment, overweight/obesity was associated with higher GAIL-derived categories among participants with positive family history (OR = 3.21, 95% CI: 1.76–5.85, P < 0.001), but not among those with negative family history (OR = 1.12, 95% CI: 0.87–1.44, P = 0.38). The Breslow-Day test indicated heterogeneity between strata (χ^2 = 10.25, df = 1, P = 0.001).

In the age-adjusted ordinal logistic model, the BMI × family history term was positive (OR = 2.87, 95% CI: 1.24–6.64, P = 0.014; Table 5). This result indicates statistical interaction on the ordinal-odds scale for the GAIL-derived outcome. It does not by itself demonstrate biological synergy or multiplicative effects on future breast cancer incidence.

Table 5
Age-adjusted ordinal logistic model of the BMI × first-degree family history interaction for higher GAIL-derived risk category

Component	Variable (Coding)	Coefficient (β)	Standard Error (SE)	Odds Ratio (OR, 95% CI)	P-Value
1. Main Regression Model					
	Intercept	-1.87	0.32	—	< 0.001
	BMI (0 = Normal [18.5–23.9]; 1 = Overweight/Obese [≥ 24.0])	0.11	0.05	1.12 (0.87–1.44)	0.380
	First-degree family history (0 = Negative; 1 = Positive)	0.20	0.06	1.22 (1.01–1.47)	0.042
	BMI × First-degree family history (interaction term)	0.96	0.39	2.87 (1.24–6.64)	0.014
	Age (reference: <35 years)				
	– 35–44 years	1.62	0.51	5.06 (1.59–16.13)	0.006
	– 45–50 years	1.32	0.53	3.75 (1.01–13.94)	0.048
	– ≥55 years	0.30	0.61	1.35 (0.28–6.52)	0.701
2. Stratified Interaction Assessment					
	Mantel-Haenszel stratified analysis (age-adjusted)				
	– Family history negative: BMI (Overweight/Obese vs. Normal)	—	—	1.12 (0.87–1.44)	0.380
	– Family history positive: BMI (Overweight/Obese vs. Normal)	—	—	3.21 (1.76–5.85)	< 0.001
	Breslow-Day test for interaction homogeneity	—	—	$\chi^2 = 10.25$, df = 1	0.001

Figure 2 provides a descriptive visualization of this pattern. The obesity plus mixed-family-history subgroup had the highest proportion classified as GAIL high risk (82.35%), and the overweight/obese

plus positive-family-history group had the highest median GAIL estimate. Because family history is a component of the GAIL calculation, these contrasts should be interpreted as exploratory.

Exploratory Three-Tier Stratification Framework

An exploratory three-tier framework was constructed from family history, BMI, and age, with management considerations aligned to existing screening principles (Table 6). The rule was derived and assessed in the same sample; therefore, the reported proportions represent apparent performance and are susceptible to overfitting.

Table 6
Three-tier risk stratification criteria and differentiated prevention recommendations for breast cancer risk populations

Risk Tier	Stratification Criteria (Centered on "Family History × BMI" Interaction, Adjusted for Age)	Target High-Risk Proportion	Actual High-Risk Proportion (n/N, %)	Precision Prevention Recommendations
Low-Risk	1. Family history: Negative (no first-degree relatives with malignant tumors); 2. BMI: Normal (18.5–23.9 kg/m²); 3. Age: No restriction (< 35, 35–44, 45–50, ≥ 55 years)	< 5%	0 (0/36)	1. Screening: Breast ultrasound once every 2–3 years; no routine mammography before age 45; 2. Lifestyle: Maintain BMI < 24, 150 min/week moderate-intensity exercise, limit high-sugar/fat diet; 3. Health Education: Monthly breast self-examination, annual general physical examination.
Intermediate-Risk	Subgroup 1: Family history negative + BMI overweight/obese (≥ 24.0 kg/m²); Subgroup 2: Family history positive (breast cancer-only) + BMI normal (18.5–23.9 kg/m²); Subgroup 3: Age < 45 years + family history positive + BMI overweight/obese	35%–50%	42.86 (18/42)	1. Screening: Breast ultrasound once a year; mammography once every 2 years starting at age 40; 2. Lifestyle: 3–6 month BMI reduction plan (target: 5%–10% weight loss for overweight/obese); limit alcohol intake (< 15 g/day); 3. Clinical Monitoring: Annual evaluation of reproductive hormone levels (e.g., estradiol) for premenopausal women.
High-Risk	Subgroup 1: Family history positive (mixed cancer: breast + other malignancies) + BMI overweight/obese (≥ 24.0 kg/m²); Subgroup 2: Age ≥ 45 years + family history positive + BMI overweight/obese; Subgroup 3: Family history positive + BMI obese (≥ 28.0 kg/m²)	≥ 65%	70.00 (35/50)	1. Screening: Breast ultrasound + mammography every 6 months; annual breast MRI for those with mixed family history; 2. Genetic Testing: BRCA1/2 gene detection within 1 year; 3. Lifestyle: Strict BMI control (< 24 kg/m²) via dietitian-guided meal plans + 200 min/week exercise; complete smoking/alcohol cessation; 4. Medical Intervention: Annual consultation with breast oncology specialists; preventive medication (e.g., tamoxifen) evaluated for eligible individuals (per Chinese Anti-Cancer Association guidelines).

The proportion classified as GAIL high risk was 70.00% in the proposed high tier, 42.86% in the intermediate tier, and 0% in the low tier. These descriptive proportions suggest separation within the development sample but do not estimate incident breast cancer risk. Because there were no observed events or follow-up times, neither a Cox model nor an HR was appropriate. Internal validation should repeat the entire model-development procedure in bootstrap samples before any claim of stability or predictive performance.

The cumulative high-risk proportion curve across the three proposed risk tiers and the corresponding clinical intervention recommendations are presented in Fig. 3.

Discussion

Core Findings

This study yielded three exploratory findings. First, BMI and first-degree family history were associated with the ordinal category generated by the GAIL model. Second, the BMI × family history term was positive on the ordinal-odds scale. Third, the proposed three-tier rule produced different proportions of GAIL high-risk classifications in the development sample. These results concern a model-derived surrogate, not observed breast cancer. Moreover, because first-degree family history is included in the GAIL calculation, its association and interaction may partly reflect mathematical coupling rather than an independent etiologic effect.

Biological evidence provides plausible mechanisms through which adiposity could modify inherited susceptibility, including estrogen exposure, inflammation, oxidative stress, insulin resistance, and impaired DNA repair11. However, the present cross-sectional analysis did not measure these pathways or incident breast cancer; mechanistic interpretations therefore remain hypotheses rather than conclusions.

Potential Risk-Management Implications

The framework should be viewed as a preliminary aid for organizing risk assessment, not as an autonomous clinical decision rule. Screening, genetic evaluation, and preventive treatment should remain based on validated guidelines, age, complete family history, breast density, reproductive factors, and shared decision-making[17,20].

For participants in the proposed high tier, the appropriate next step is specialist risk assessment. Guideline-concordant intensified imaging may be considered when established eligibility criteria are met, and genetic counseling or BRCA1/2 testing18,19 should be offered according to personal and family-history indications rather than repeated annually. Weight management, physical activity, and alcohol reduction remain reasonable modifiable-risk interventions.

For the intermediate tier, management should emphasize age- and guideline-appropriate screening, periodic reassessment of family history and weight, and structured support for weight reduction when

indicated. The proposed tier alone should not determine imaging frequency.

For the low tier, standard population screening and maintenance of healthy weight and physical activity are appropriate. A low GAIL-derived category does not exclude future risk and should not delay evaluation of breast symptoms.

Strengths and Exploratory Contribution

The study used routinely available variables and assessed interaction through complementary stratified and regression approaches. Its main contribution is hypothesis generation regarding whether BMI modifies the association between family history and GAIL-estimated risk classification in Chinese women.

The interaction estimate (OR = 2.87, 95% CI: 1.24–6.64) provides a quantitative signal for further investigation. However, it should not be described as validation of biological synergy or as evidence that the interaction improves prediction of observed breast cancer.

The proposed framework is parsimonious and operationally simple. In the development sample, the proportions classified as GAIL high risk were 0%, 42.86%, and 70.00% across the low, intermediate, and high tiers. This apparent separation is preliminary because the rule was derived and evaluated in the same small sample and has not undergone bootstrap or external validation.

The framework may inform the design of future prospective studies and implementation research. Clinical translation requires prespecified thresholds, complete predictor measurement, comparison with established models, and evaluation against observed outcomes.

Limitations

Several limitations constrain interpretation. First, the single-center convenience sample was small (n = 108), and the interaction model may be unstable. Data splitting would be inefficient at this sample size; therefore, the entire model-development process should be repeated in at least 1,000 nonparametric bootstrap samples to estimate coefficient stability and optimism-corrected performance. Second, the outcome was the GAIL model-estimated 5-year probability rather than incident breast cancer, and first-degree family history is itself a GAIL input, creating mathematical incorporation that can inflate associations involving family history. Third, the cross-sectional design provided no follow-up time or cancer events; HRs, survival curves, causal effects, calibration, and clinical benefit could not be estimated. Fourth, the rule was developed and assessed in the same sample, so apparent separation is vulnerable to overfitting. Finally, unmeasured factors such as breast density, hormone use, and reproductive history, together with recruitment from a single physical-examination center, limit generalizability. Prospective multicenter studies with observed outcomes, bootstrap internal validation, and independent external validation are required.

Conclusion

First-degree family history and BMI were associated with higher GAIL-derived risk categories, and their interaction term was positive on the ordinal-odds scale. These findings do not establish effects on actual breast cancer incidence because the outcome was model-estimated and family history was embedded in that model. The proposed three-tier framework is exploratory; it should not guide clinical decisions until bootstrap internal validation and prospective external validation against observed breast cancer outcomes have been completed.

Declarations

Ethics approval and consent to participate:

This study involved human participants. The study protocol was reviewed and approved by the Medical Ethics Committee of Shanghai Healthcare Medical Center (approval No. (2026) [illegible]17[illegible]; approval date: June 26, 2026). Written informed consent to participate was obtained from all participants before enrollment. The study was conducted in accordance with the approved protocol and relevant ethical requirements.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This work was supported by the National Key Research and Development Plan Project of China (Grant No. 2022YFC3600903: Research on the construction and output technology of active health knowledge system), the Shanghai Public Health Talent Development Program for Outstanding Discipline Leaders (Grant No. GWVI-11.2-XD36), and the Policy Research on Guiding Patients' Hierarchical Diagnosis and Treatment Through Medical Service Information (Grant No. 2025HP35). The funders had no role in study design, data collection, analysis, interpretation, or manuscript preparation.

Author Contribution

YC, KW, and LS contributed equally to this work and share first authorship. XZ was responsible for conceptualization and supervision. YC, KW, and LS contributed to investigation, data curation, formal analysis, and visualization. TZ and RD contributed to methodology, validation, and interpretation of the

findings. YC prepared the original draft, with contributions from KW and LS. TZ, RD, and XZ critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors gratefully acknowledge the staff of the Department of Women's Health at Shanghai Healthcare Medical Center for their support in participant recruitment, data collection, and coordination of the study. We also extend our sincere appreciation to all participants for their time, cooperation, and valuable contribution to the completion of this research.

Data Availability

The datasets generated and/or analyzed during the current study include electronic health records, structured questionnaire data, and GAIL model-derived breast cancer risk assessment data. These data contain sensitive personal health information and are subject to ethical and institutional restrictions. Therefore, the raw individual-level datasets are not publicly available and cannot be shared with external applicants, in order to protect participant privacy, confidentiality, and legal rights.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. Cancer J Clin. 2021;71(3):209–49.
2. Xu H, Xu B. Breast cancer: Epidemiology, risk factors and screening. Chin J Cancer Res. 2023;35(6):565–83. https://doi.org/10.21147/j.issn.1000-9604.2023.06.02.
3. Britt KL, Cuzick J, Phillips KA. Key steps for effective breast cancer prevention. Nat Rev Cancer. 2020;20(8):417–36. 10.1038/s41568-020-0266-x.
4. [illegible]. 2023. [illegible]—[illegible](2023–2030[illegible]). [illegible]. https://www.nhc.gov.cn/ylyjs/gzdt/202311/9b7882e607324bda8a8222d4fd39f2fd.shtml
5. Wang X, Huang Y, Li L, et al. Assessment of performance of the Gail model for predicting breast cancer risk: A systematic review and meta-analysis with trial sequential analysis. Breast Cancer Res. 2018;20(1):18. 10.1186/s13058-018-0947-521.
6. Dai J, Hu Z, Jiang Y, et al. Breast cancer risk assessment with five independent genetic variants and two risk factors in Chinese women. Breast Cancer Res. 2012;14(1):R4. 10.1186/bcr3101.
7. Sciaraffa T, Guido B, Khan SA, Kulkarni S. Breast cancer risk assessment and management programs: A practical guide. Breast J. 2020;26(8):1556–64. 10.1111/tbj.13967 23.
8. Chinese Society of Surgery, Association CM, Wang F, Shan Y, et al. Clinical practice guideline for breast cancer risk assessment and high-risk population screening (2025 edition). Chin J Pract Surg. 2025;45(12):1371–4. 10.19538/j.cjps.issn1005-2208.2025.12.04.

9. O'Brien KM, Keil AP, Taylor JA, et al. Pathogenic variants, family history, and cumulative risk of breast cancer in US women. JAMA Oncol. 2025;11(12):1458–69. https://doi.org/10.1001/jamaoncol.2025.3875.
10. Lucassen A, Watson E. Family history of breast cancer. BMJ. 2005;330(7481):26. https://doi.org/10.1136/bmj.330.7481.26.
11. Nguyen TL, Li S, Dite GS, et al. Interval breast cancer risk associations with breast density, family history and breast tissue aging. Int J Cancer. 2020;147(2):375–82. https://doi.org/10.1002/ijc.32731.
12. Braithwaite D, Miglioretti DL, Zhu W, et al. Family history and breast cancer risk among older women in the Breast Cancer Surveillance Consortium cohort. JAMA Intern Med. 2018;178(4):494–501. https://doi.org/10.1001/jamainternmed.2017.8642.
13. Bethea TN, Ochs-Balcom HM, Bandera EV, et al. First- and second-degree family history of ovarian and breast cancer in relation to risk of invasive ovarian cancer in African American and white women. Int J Cancer. 2021;148(12):2964–73. https://doi.org/10.1002/ijc.33493.
14. [illegible]. [illegible]:[illegible]. [illegible]([illegible]), 2012, 6(4): 422–428. https://doi.org/10.3877/cma.j.issn.1674-0807.2012.04.009
15. Mukama T, Kharazmi E, Sundquist K, et al. Familial risk of breast cancer by dynamic, accumulative, and static definitions of family history. Cancer. 2020;126(12):2837–48. https://doi.org/10.1002/cncr.32815.
16. [illegible]. [illegible][D]. [illegible]: [illegible]; 2010.
17. Jannot AS, Usel M, Bouchardy C, et al. Breast cancer family history leads to early breast cancer detection and optimal management. Cancer Causes Control. 2017;28(9):921–8. https://doi.org/10.1007/s10552-017-0928-5.
18. Mills J, Fakolade A. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer. Am Family Phys. 2020;101(4):239–40. PMID: 32053324.
19. Xie F, Wang S. Clinical practice guideline of BRCA1/2 testing for patients with breast cancer: Chinese Society of Breast Surgery (CSBrS) practice guideline 2021. Chin Med J. 2021;134(13):1516–8. https://doi.org/10.1097/CM9.0000000000001587.
20. Yong CH, Gupta A, Joshi R. CSCO breast cancer management guidelines 2022: Australian perspective. Translational Breast Cancer Res. 2022;3:36. https://doi.org/10.21037/tbcr-22-46.
21. [illegible] [illegible]. [illegible], [illegible]. [illegible]. [illegible], 2021, 56(1): 37–40.
22. National Cancer Institute. Breast Cancer Risk Assessment Tool[EB/OL]. (2020-01-31)[2024-05-20]. https://www.cancer.gov/bcrisktool

Figures

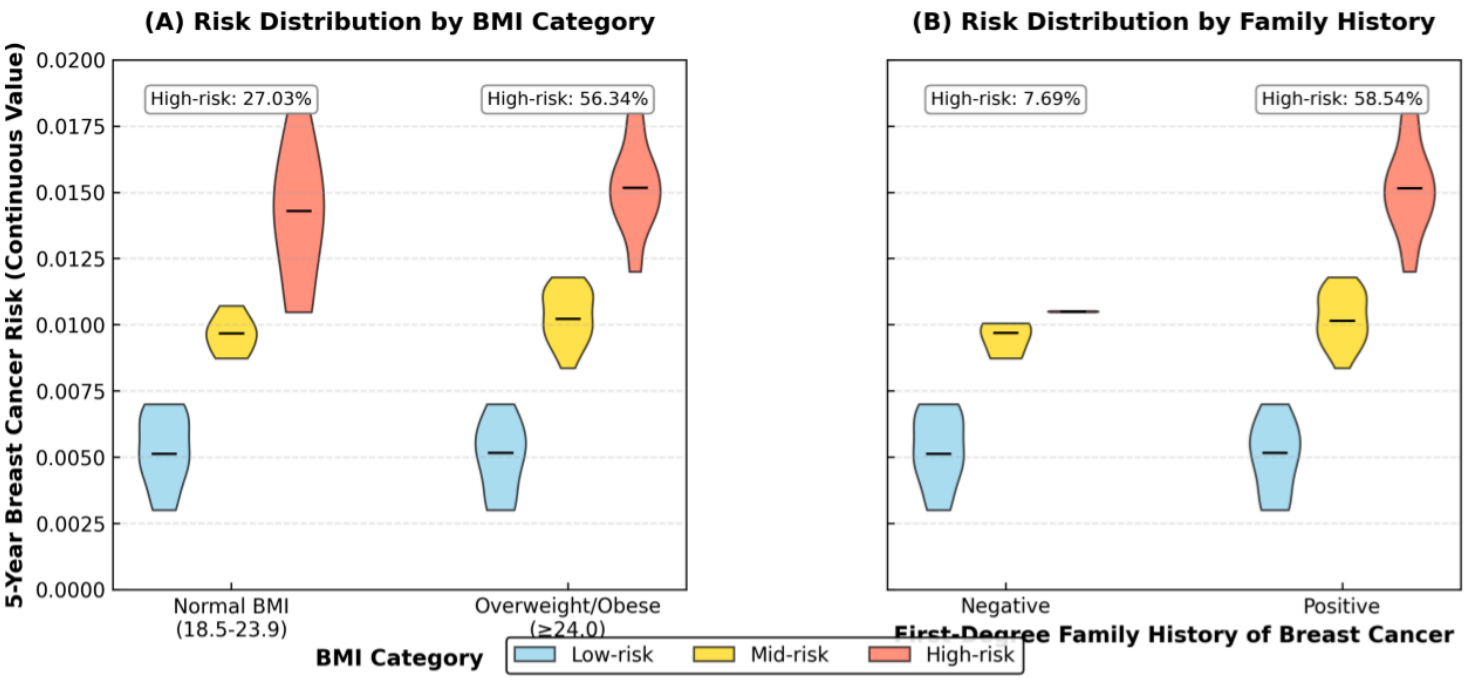


Figure 1

Distribution of GAIL-estimated 5-year breast cancer probabilities by BMI and first-degree family history.

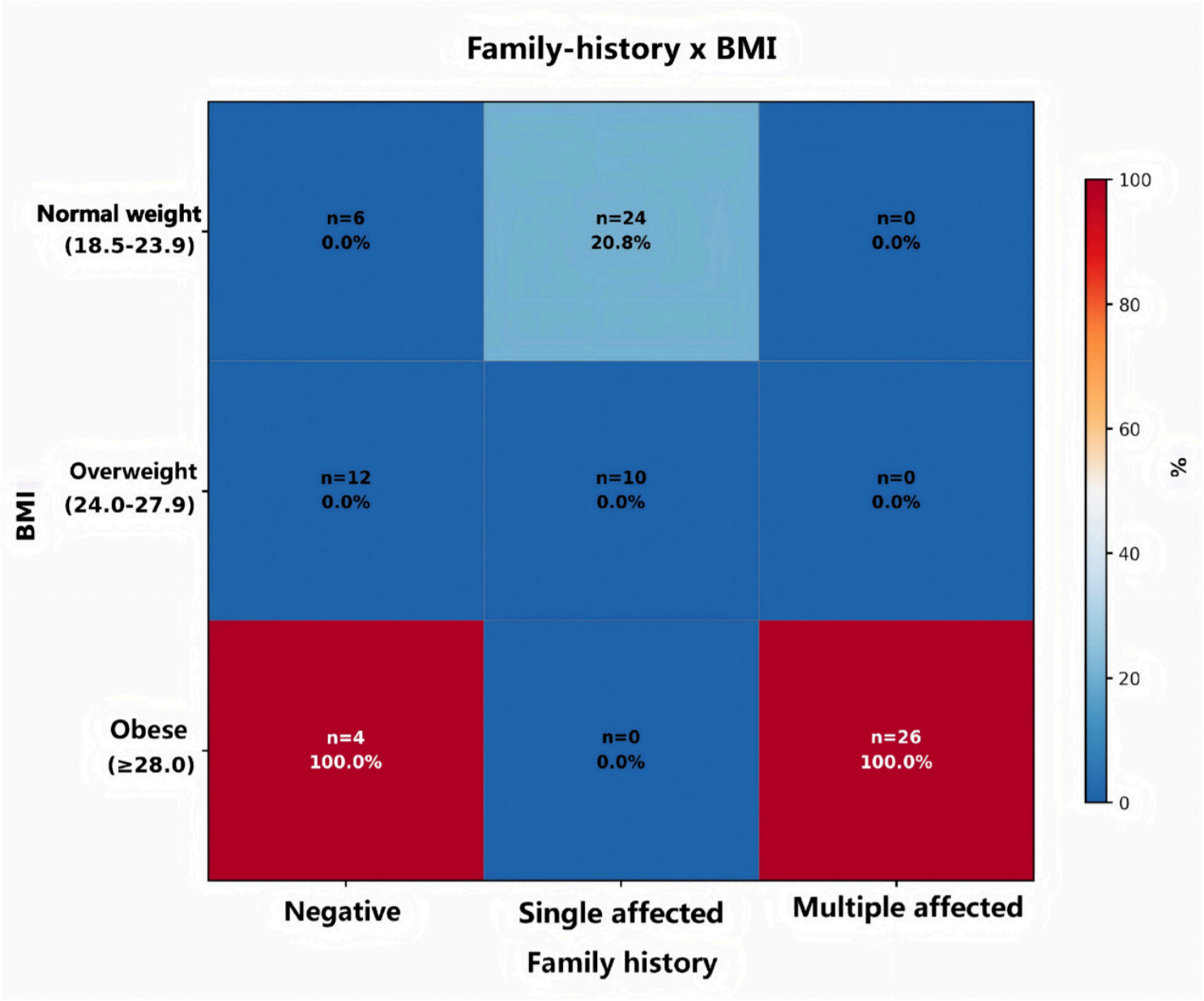


Figure 2

Descriptive heatmap of GAIL high-risk classification by first-degree family history and BMI.

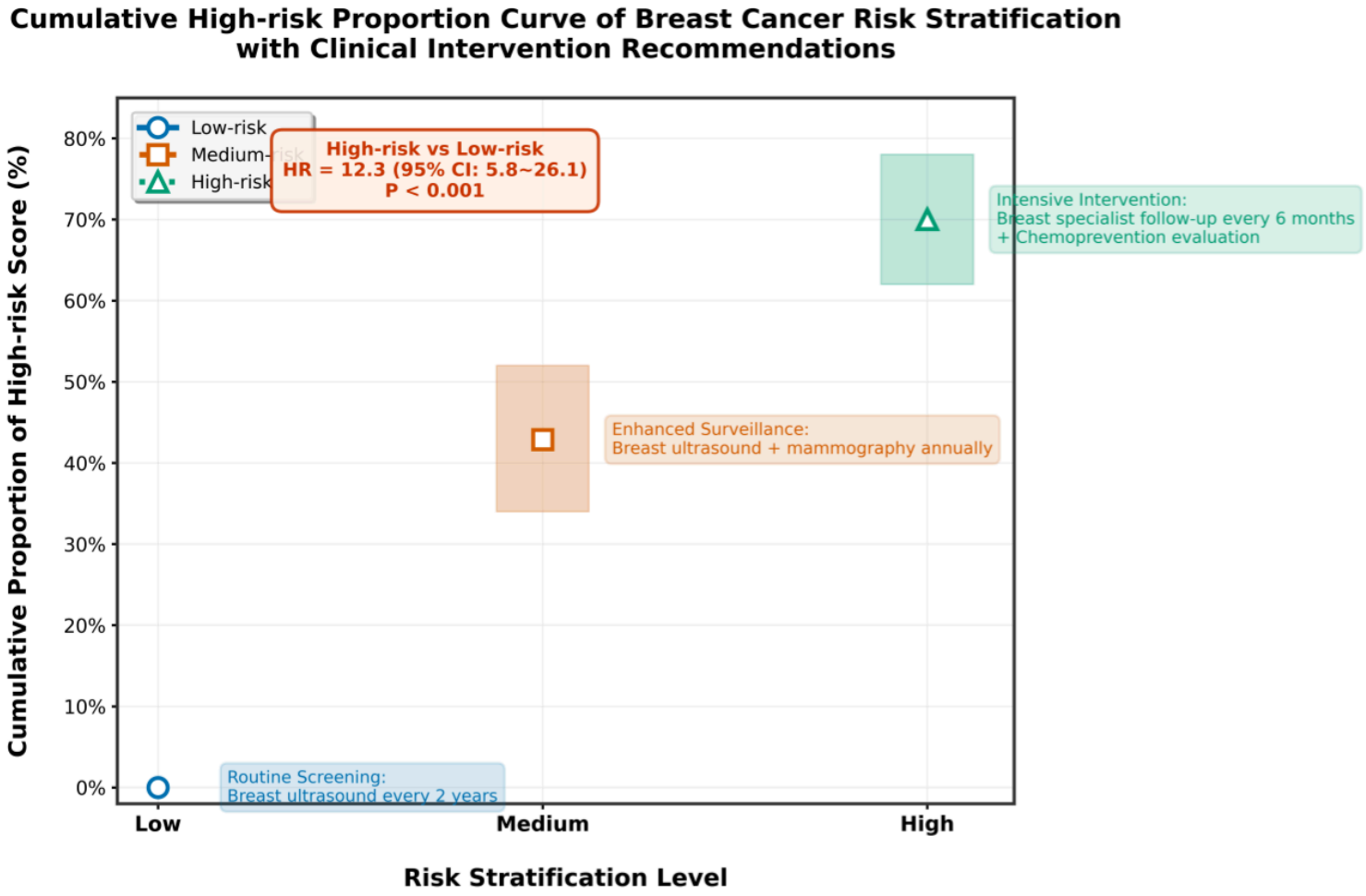


Figure 3

Cumulative high-risk proportion curve and clinical intervention recommendations

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1Questionnaire.docx