Wang *et al. BMC Cancer* (2025) 25:1234
https://doi.org/10.1186/s12885-025-14651-6

BMC Cancer

RESEARCH **Open Access**

The application and predictive value of the weight-adjusted-waist index in BC prevalence assessment: a comprehensive statistical and machine learning analysis using NHANES data

Wenjing Wang[1], Biao Wu[1], Jian Li[1], Yibiao Shang[1], Mengting Liu[1], Qi Fang[1], Han Zhang[1], Xiang Li[1] and Dongdi Wu[1*]

Abstract

Background Obesity is a known risk factor for breast cancer (BC), but conventional metrics such as body mass index (BMI) may insufficiently capture central adiposity. The weight-adjusted waist index (WWI) has emerged as a potentially superior anthropometric marker of central adiposity, as it provides a more accurate reflection of fat distribution around the abdomen compared to traditional measures such as BMI. This study aimed to investigate the association between WWI and BC prevalence using data from a nationally representative population in the United States.

Methods A total of 10,760 women aged over 20 years from the 2005–2018 National Health and Nutrition Examination Survey were included. Logistic regression was used to assess the association between WWI and BC prevalence. Multicollinearity was addressed using variance inflation factor diagnostics. Machine learning methods, including random forest and LASSO regression, were employed for variable selection and model comparison. The performance of the models was evaluated using ROC curves, calibration plots, and decision curve analysis.

Results In unadjusted models, WWI was significantly associated with BC (odds ratio (OR) = 1.56; 95% confidence interval (CI): 1.32–1.86). However, in the fully adjusted model, the association with BC was no longer statistically significant (OR = 0.98; 95% CI: 0.75–1.26). Machine learning models ranked WWI as one of the top predictors, with the random forest model retaining WWI as an important variable, while LASSO excluded it. Models based on variables selected by both LASSO and random forest, which included WWI, were built and assessed using ROC curve analysis. The random forest and LASSO models achieved AUCs of 0.795 and 0.79, respectively, demonstrating improved predictive performance. These findings suggest that while WWI may not serve as an independent predictor of BC, it may offer additional value when combined with other key covariates.

Conclusion Although the WWI was related to BC prevalence before multivariable adjustment, it was not significantly linked to BC after adjustment. Given the cross-sectional design and the relatively small sample of BC cases ($n = 326$),

---

*Correspondence:
Dongdi Wu
yfywdd31@163.com

Full list of author information is available at the end of the article

the findings should be viewed with caution. Future research with larger prospective cohorts is needed to confirm these results and explore WWI's role in BC risk stratification. Studies should also investigate whether WWI can serve as a reliable independent predictor of BC in future research, taking into account other factors that may influence the association.

Keywords National health and nutrition examination survey, Breast cancer, Weight-adjusted-waist index, Obesity, Central adiposity, Machine Learning

Background

Breast cancer (BC) is the most diagnosed cancer among women worldwide [1]. The incidence of BC has shown a consistent upward trend. In 2020, BC became the most frequently diagnosed cancer worldwide, with over 2.3 million new cases and approximately 685,000 deaths; by 2040, the global burden is projected to exceed 3 million new cases and 1 million deaths annually, with developing regions bearing a disproportionate share of mortality despite higher incidence in developed countries [2].

Obesity is a well-established and prevalent risk factor for BC. Obesity, now ranked as the fifth leading cause of death worldwide [3], has been linked to an increase in obesity-related diseases, including BC. Obesity contributes to the development of various chronic diseases such as asthma, diabetes, cardiovascular diseases, and cancer [4]. Furthermore, obesity promotes tumor development through mechanisms such as local fat inflammation and alterations in the tumor microenvironment [5]. Adipocytes play a significant role in the onset and progression of BC [5–7]. Body mass index (BMI) is the most commonly used measure of obesity. According to the World Health Organization (WHO), overweight is defined as a BMI of 25-<30 kg/m^2 and obesity as a BMI > 30 kg/m^2 in general populations, while for Asian populations, lower thresholds (BMI ≥ 23 for overweight and ≥ 27.5 for obesity) are recommended due to higher metabolic risk at lower BMI levels [8]. However, BMI has limitations as an indicator of obesity, as it cannot distinguish between fat and lean body mass and does not directly reflect obesity status [9]. Additionally, BMI's effectiveness in representing obesity may vary across different racial groups [10], and the"obesity paradox"suggests an inconsistent or even inverse correlation between BMI and mortality [11]. Thus, BMI alone is insufficient for assessing the impact of obesity on BC, and more accurate and stable assessment tools are needed to better define obesity and identify high-risk patients for cancer development [5].

In 2018, the weight-adjusted waist index (WWI; waist circumference/√weight) was introduced as a novel anthropometric measure that more specifically captures central adiposity [11]. WWI may be a more reliable predictor of obesity-related health risks because central adiposity, which it captures better than BMI, has been more strongly linked to such risks in prior studies [12]. While BMI reflects general adiposity, WWI's focus on central adiposity may be particularly relevant for evaluating health risks, including the risk of BC [11]. While extensive research exists on the link between obesity and BC, studies specifically examining the relationship between central adiposity as measured by WWI and BC are still limited.

In the present study, we evaluate the association between WWI and BC prevalence in a nationally representative sample of adult US women, using advanced statistical methods to adjust for potential confounders and assess the predictive value of WWI for BC outcomes. Previous research has explored the association between WWI and BC prevalence among adult women using NHANES (National Health and Nutrition Examination Survey) data from 2011 to 2018 [13]. The present study builds on this work by analyzing a broader time frame (2005–2018), thereby increasing statistical power and generalizability. Additionally, the covariates considered in this study differ from those included in the previous research, providing further insights into the relationship between WWI and BC prevalence.

Research methodology

NHANES, initiated by the United States National Center for Health Statistics, is a national cross-sectional study that evaluates the well-being and dietary conditions of American citizens every two years, aiming to gather thorough insight into current illness patterns and offer guidance for creating public health strategies. The NHANES webpage was visited at https://www.cdc.gov/nchs/nhanes/ to access all of the datasets.

This study analyzed data from three consecutive NHANES cycles between 2005 and 2018, with a total sample of 67,203 female participants. Participants who lacked BC outcome data ($n = 27,457$) and had missing information at baseline ($n = 11,455$), and individuals aged < 20 years, who were excluded during the data merging process, were excluded from the analysis. Additionally, participants with missing or non-informative responses (e.g.,'Don't know','Refused') for covariate data were excluded, yielding a final analytic sample of 10,760 individuals. The participant selection process is illustrated in Fig. 1.

To address concerns regarding missing data, we performed Little's MCAR test to assess the missingness mechanism. The test results showed that the missing data

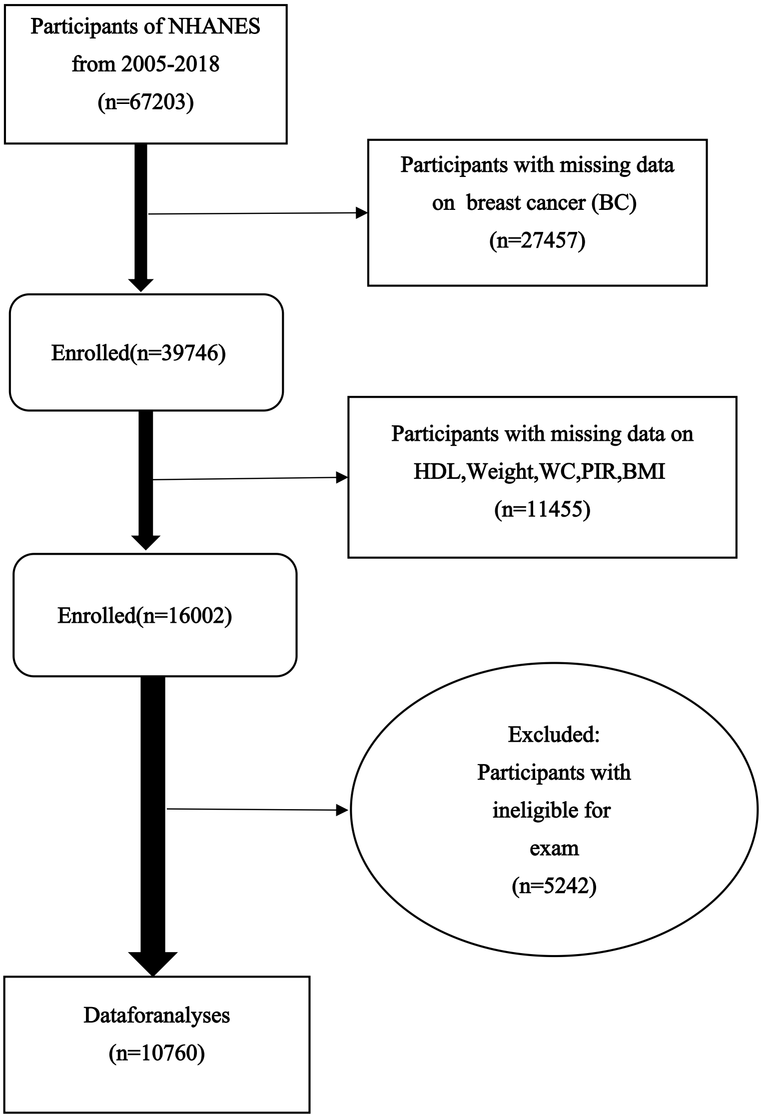


Fig. 1 Flowchart of participant selection from NHANES 2005–2018

were not missing completely at random. To minimize the potential bias introduced by non-random missingness, we used a weighting procedure based on NHANES standard methods. This procedure applies population weights such as the interview weight and examination weight, which adjust for selection probabilities, nonresponse, and the complex survey design. These weights help ensure the sample is representative of the U.S. population and mitigate the impact of missing data on the validity and generalizability of the results.

Diagnosis criteria for BC

The Medical Condition Questionnaire serves as the source for self-reported BC diagnosis. Participants were asked whether they had ever been informed by a doctor or other healthcare professionals of having cancer or any type of malignant tumor. If the response was"yes", they were further asked,"What type of malignant tumor was it?"Those who self-reported a diagnosis of BC were classified as BC cases, while those not diagnosed with malignant tumors or diagnosed with malignancies other than BC were categorized as non-cases.

Exposure variable

WWI, the main exposure variable in this study, was calculated as waist circumference (cm) divided by the square root of weight (kg), i.e., WWI = waist circumference/√weight. WWI was analyzed both as a continuous variable (per 1-unit increment) and as a categorical variable, stratified into quartiles based on the weighted distribution of the study population. To assess linear trends across quartiles, the median value of each group was assigned and treated as a continuous variable in the regression models to compute the P for trend.

Covariates

Inclusion criteria included having complete data on all the following covariates: age, race, education level (EA), marital status (Mar.Status), and poverty-income ratio (PIR), BMI, Total cholesterol (TC), high-density lipoprotein (HDL), smoking status, Hypertension (HTN), waist circumference (WC), weight, menopausal (MEN) and Gravida. All covariates, except for TC, HDL, WC, and weight, were self-reported through questionnaires administered by NHANES personnel. Smoking status was based on whether the participant had smoked at least 100 cigarettes in their lifetime (yes/no). Hypertension was defined as a self-reported prior diagnosis of high blood pressure by a health professional (yes/no). Menopause is defined as having had at least one menstrual period in the past 12 months, excluding bleeding caused by medical conditions, hormone therapy, or surgeries. If a woman has had a menstrual period within the past 12 months, she is considered premenopausal; if she has not had a menstrual period in the past 12 months, she is considered postmenopausal.

WC and weight were measured by trained health technicians in NHANES mobile examination centers using standardized protocols. Quality assurance procedures were implemented to ensure accuracy and consistency across measurement sites. Total cholesterol (TC) and high-density lipoprotein (HDL) levels were measured from blood samples collected during the physical examination. Venous blood was drawn by certified phlebotomists in mobile examination centers using standardized protocols. Laboratory analyses were conducted in NHANES-certified laboratories following rigorous quality control procedures to ensure measurement accuracy.

The levels of categorical variables were defined as follows:

Sex: male/female; Race: Mexican American/Other Hispanic/Non-Hispanic White/Non-Hispanic Black/Other race (including multiracial);Education level: less than 9th grade/9–11th grade (including 12th grade, no diploma)/high school graduate or GED/some college or associate degree/college graduate or above; Marital status: married/widowed/divorced/separated/never married/living with partner; Smoking, Hypertension: yes/no, Menopausal: pre – menopausal/post – menopausal. All variable codings followed NHANES definitions.

WC and weight were initially considered as covariates but excluded from the final models due to strong collinearity with WWI.

Statistical assessments

All analyses incorporated the complex, multistage sampling design of NHANES. Sampling weights (WTMEC2YR) were applied to ensure estimates were representative of the U.S. civilian non-institutionalized population. Variance estimation accounted for stratification and clustering using SDMVSTRA and SDMVPSU. Weighted means, proportions, and regression models were calculated in accordance with NHANES analytic guidelines.

All statistical analyses were performed using R (version 4.4.3). The following R packages were used:dplyr, tidyverse, plyr, readr, stringr, tidyr, arsenal, gtsummary, tableone, gtsummary, forestplot, forestploter, survey, survival, survminer, Hmisc, splines, riskRegression, pROC, ggplot2, plotly, ggthemes, ggsci, ggpubr, ggsurvfit, regplot, patchwork, gridExtra, scales, rms, randomForest, caret, glmnet, Matrix, randomForestExplainer, vip, car, foreign, jskm, devtools, remotes, contsurvplot, shape, glmnet, reshape2, tidymodels, modeldata, recipes, rsample, grid, readx, VIM, naniarl. The default settings of each package were applied, with no modifications made.

Categorical variables were described using counts and percentages, and comparisons between groups were made via the Pearson chi-square test. For continuous variables, medians and interquartile ranges (IQRs) were reported, and the Kruskal–Wallis rank-sum test was applied to assess differences across groups.

To investigate the relationship between the weight-adjusted WWI and BC prevalence, multivariable logistic regression models were built. During model development, multicollinearity among covariates was assessed using generalized variance inflation factors (GVIFs). A GVIF greater than 10 was considered indicative of significant collinearity. Variables with high GVIFs were excluded prior to constructing the final regression models, ensuring better model stability and interpretability.

TC, HDL, and hypertension were included in the multivariable models as covariates based on their established associations with both abdominal adiposity and BC prevalence. Although these variables may be influenced by adiposity, they were treated as potential confounders to isolate the direct association between WWI and breast cancer. This approach reflects a conservative modeling strategy and may underestimate the total effect of WWI. The possibility that these factors act as mediators is acknowledged as a limitation of the present analysis.

To minimize the risk of overfitting, the number of covariates included in the final logistic regression model was limited to 13. Given the 326 BC events observed, the events-per-variable (EPV) ratio was approximately 25.08, which exceeds the commonly accepted threshold of 10 EPV for reliable model estimation.

We also explored potential non-linear associations between WWI and both BC prevalence and prognosis using restricted cubic spline (RCS) models. Subgroup analyses were carried out to assess whether metabolic status, comorbid conditions, or lifestyle factors modified these associations.

In addition to traditional regression models, we applied LASSO (Least Absolute Shrinkage and Selection Operator) logistic regression to select key variables associated with BC. The optimal value of the penalty parameter (λ) was chosen based on ten-fold cross-validation, aiming to minimize the mean squared error.To further assess variable importance and model performance, we used a random forest algorithm implemented in the randomForest package in R. The model was built with 500 trees, and variable importance was evaluated using the mean decrease in Gini index and minimal depth analysis.

ROC curves were constructed for all models to evaluate discriminative performance, while calibration was assessed using bootstrap-based calibration curves. Decision curve analysis (DCA) was conducted to estimate net clinical benefit across a range of threshold probabilities, comparing the final model against treat-all and treat-none strategies.

Variables identified by Random Forest and LASSO were assessed for multicollinearity using the GVIFs. Weight and WC were excluded, while WWI was retained for its clinical relevance.

After removing collinear variables, separate logistic regression models were constructed based on the remaining variables from the Random Forest and LASSO selections, respectively, to evaluate associations with BC.

Sensitivity analyses

To assess the robustness of our findings, two sensitivity analyses were conducted. First, to examine whether the inclusion of BMI influenced the association between WWI and BC, BMI was added to the primary multivariable logistic regression model. Second, we replicated the model specification used in Huang et al. [13], adjusting for the same set of covariates: age, race, education level, marital status, PIR, total cholesterol, alcohol consumption, smoking status, hypertension, diabetes mellitus, coronary heart disease, and physical activity. This analysis was performed on the same analytic sample (NHANES 2005–2018) as used in the main model. Statistical significance was defined as $P < 0.05$.

Result

Baseline traits of the investigated population

Table 1 presents the baseline demographic and clinical characteristics stratified by BC status. Significant differences were observed between participants who reported a diagnosis of BC and those who did not, in terms of age ($P < 0.001$), race ($P < 0.001$), PIR ($P = 0.034$), marital status ($P < 0.001$), hypertension ($P < 0.001$), WC ($P = 0.026$), menopausal ($P < 0.001$), and WWI ($P < 0.001$).

Participants who reported a diagnosis of BC may have been older, had a higher incidence of comorbidities (such as hypertension), higher total cholesterol levels, higher obesity indices, and may have been more likely to be postmenopausal. High WWI and WC were especially prevalent among BC cases compared to non-cases, whereas differences in BMI prevalence were less pronounced.

Participants were divided into quartiles based on their WWI values in Table S1 (Q1: 8.59–10.48,Q2: 10.48–11.03,Q3: 11.03–11.55,Q4: 11.55–15.39). As the WWI value increased, the prevalence of BC steadily rose (Q1: 1.7%, Q2: 3.2%, Q3: 3.3%, Q4: 4.6%, $P < 0.001$). Significant differences were observed across all domains in the WWI quartiles ($P < 0.001$). Compared to the first quartile, participants in the fourth quartile were older, had lower education levels, worse economic status, a higher prevalence of hypertension, and a greater likelihood of smoking and being postmenopausal.

Association between WWI and BC prevalence

Table 2 presents the association between WWI and BC based on logistic regression analyses. In the unadjusted model, a significant positive association was observed when WWI was analyzed both as a continuous variable (OR = 1.56, 95% CI: 1.32–1.86, $P < 0.001$) and by quartiles (P for trend < 0.001). Prior to model construction, multicollinearity diagnostics were conducted for all covariates (Table S2). Weight and waist circumference (WC) were excluded due to high GVIFs of 142.676 and 250.284, respectively. BMI had a moderate GVIF of 7.056, indicating a modest correlation with WWI. We then conducted a sensitivity analysis to assess the relationship between WWI and BC, with and without BMI as a covariate (Table S3). The results showed that the exclusion of BMI did not substantially change the effect estimate for WWI on BC: the odds ratio was 0.98 (95% CI: 0.75–1.26; $p = 0.900$) in the model with BMI and 0.99 (95% CI: 0.77–1.27; $p > 0.900$) in the model without BMI. Despite this, BMI was included in the final model to account for its potential confounding effect, given its established role as a risk factor for BC. In this model, the association between WWI and BC was no longer statistically significant, either when modeled continuously (OR = 0.98, 95% CI: 0.75–1.26, $P = 0.900$) or categorically (P for trend = 0.714).

To examine the robustness of our findings, we performed a sensitivity analysis using the same covariates as Huang et al. [13]. The association between WWI and BC remained non-significant (OR = 0.97; 95% CI: 0.75–1.24; $p = 0.800$), consistent with the primary model (OR = 0.98; 95% CI: 0.75–1.26; $p = 0.900$) (Table S8).

We applied RCS models to explore whether the relationship between WWI and BC was non-linear (Fig. 2). In both unadjusted and adjusted models, WWI showed a significant association with BC, but no evidence of a non-linear pattern was found.

Table 1 Baseline characteristics of BC group versus non-BC group

Characteristic	N[1]	Overall *n* = 64,558,024[2]	BC (No) *n* = 62,499,721[2]	BC (Yes) *n* = 2,058,303[2]	*p*-value[3]
Age	10,760	50.00 (38.00, 63.00)	50.00 (38.00, 62.00)	66.00 (59.00, 76.00)	< 0.001
Race	10,760				< 0.001
Mexican American		1,822 (8.0%)	1,788 (8.1%)	34 (3.6%)	
Other Hispanic		1,126 (5.5%)	1,101 (5.6%)	25 (2.7%)	
Non-Hispanic White		4,678 (69%)	4,483 (69%)	195 (83%)	
Non-Hispanic Black		2,131 (11%)	2,079 (11%)	52 (7.1%)	
Other Race—Including Multi-Racial		1,003 (6.6%)	983 (6.7%)	20 (3.8%)	
EA	10,760				0.091
Less Than 9th Grade		1,114 (5.3%)	1,084 (5.3%)	30 (3.8%)	
9-11th Grade (Includes 12th grade with no diploma)		1,561 (11%)	1,516 (11%)	45 (9.8%)	
High School Grad/GED or Equivalent		2,466 (23%)	2,394 (23%)	72 (23%)	
Some College or AA degree		3,393 (33%)	3,300 (33%)	93 (29%)	
College Graduate or above		2,226 (27%)	2,140 (27%)	86 (34%)	
Mar.Status	10,760				< 0.001
Married		5,657 (60%)	5,488 (60%)	169 (57%)	
Widowed		1,358 (9.9%)	1,272 (9.5%)	86 (23%)	
Divorced		1,528 (14%)	1,475 (14%)	53 (18%)	
Separated		445 (2.9%)	437 (2.9%)	8 (1.1%)	
Never married		1,007 (7.1%)	1,003 (7.4%)	4 (0.6%)	
Living with partner		765 (6.7%)	759 (6.9%)	6 (1.1%)	
PIR	10,760	2.90 (1.46, 4.99)	2.88 (1.46, 4.97)	3.21 (1.72, 5.00)	0.034
BMI	10,760	27.80 (24.30, 32.10)	27.80 (24.30, 32.10)	27.70 (24.30, 31.71)	0.836
TC	10,760	198.00 (172.00, 226.00)	197.00 (172.00, 225.00)	202.00 (177.00, 231.00)	0.104
Smoking	10,760				0.293
Yes		4,055 (41%)	3,931 (41%)	124 (37%)	
No		6,705 (59%)	6,503 (59%)	202 (63%)	
HTN	10,760				< 0.001
Yes		4,061 (34%)	3,865 (34%)	196 (53%)	
No		6,699 (66%)	6,569 (66%)	130 (47%)	
HDL	10,760	57.00 (47.00, 68.00)	57.00 (47.00, 68.00)	58.00 (48.00, 71.00)	0.405
WC	10,760	95.50 (86.50, 105.10)	95.40 (86.40, 105.20)	97.10 (89.00, 104.20)	0.022
Weight	10,760	72.60 (63.50, 84.10)	72.60 (63.50, 84.20)	71.70 (64.80, 81.60)	0.497
Menopausal	10,760				< 0.001
Pre – menopausal		4,672 (44%)	4,654 (45%)	18 (6.8%)	
Post – menopausal		6,088 (56%)	5,780 (55%)	308 (93%)	
Gravida	10,760	3.00 (2.00, 4.00)	3.00 (2.00, 4.00)	3.00 (2.00, 4.00)	0.442
WWI	10,760	11.14 (10.64, 11.69)	11.14 (10.63, 11.68)	11.43 (10.87, 12.06)	< 0.001

Variable definitions are consistent with those described in the Methods

Abbreviations: *EA* educational attainment, *Mar.Status* marital status, *PIR* poverty-to-income ratio, *BMI* body mass index, *TC* total cholesterol, *WC* waist circumference, *HTN* hypertension, *HDL* high-density lipoprotein, *BC* breast cancer, *WWI* weight-adjusted-waist index

[1] N not Missing (unweighted)

[2] n (weighted) (%); Median (Q1, Q3) (weighted using NHANES sampling weights to reflect the U.S. civilian non-institutionalized population)

[3] Design-based KruskalWallis test; Pearson's X^2: Rao & Scott adjustment

Subgroup analysis of the relationship between WWI and BC prevalence

Subgroup analysis showed that the association between WWI and BC did not differ significantly across the subgroups (Fig. 3; all P for interaction > 0.05), suggesting that this association remained generally stable across all subgroups.

Machine learning

Variable importance and interaction effects in random forest modeling

Random forest analysis identified WWI as the top predictors of BC diagnosis, based on variable importance measures including Gini index, accuracy-based ranking, and default model contribution (Fig. 4). The WWI showed a minimal depth of 3.04 in the decision tree model, indicating its significant role in predicting breast cancer risk,

Table 2 Association between WWI and BC prevalence in logistic regression models

	Crude model OR (95% CI) *P*-value	**Model 1 OR (95% CI) *P*-value**	**Model 2 OR (95% CI) *P*-value**
WWI	1.56 (1.32,1.86) < 0.001	1.04 (0.83,1.29) 0.700	0.98 (0.75,1.26) 0.900
WWI quartiles			
Q1	Ref	Ref	Ref
Q2	1.90 (1.12,3.23) 0.017	1.35 (0.79,2.31) 0.300	1.21 (0.69,2.11) 0.500
Q3	2.02 (1.30,3.12) 0.002	1.12 (0.71,1.77) 0.600	0.95 (0.58,1.55) 0.800
Q4	2.81 (1.79,4.41) < 0.001	1.15 (0.70,1.89) 0.600	0.97 (0.54,1.74) > 0.900
P for trend	< 0.001	0.769	0.714

Model 1: adjust age, race, poverty-to-income ratio, marital status, educational attainment

Model 2: adjust age, race, poverty-to-income ratio,marital status,educational attainment, body mass index, total cholesterol, smoke, hypertension, high-density lipoprotein, gravida, menopausal

WWI quartiles were defined as follows: Q1 (<10.48), Q2 (10.48–11.03), Q3 (11.03–11.55), Q4 (>11.55)

albeit at a relatively deeper level in the tree (Fig. 5). In the multi-way importance plot, WWI showed strong contributions to both prediction accuracy and node purity (Fig. 6). Interaction analysis identified frequent pairings between age, WWI, BMI, WC, weight, menopausal status, and lipid variables such as HDL and TC. These interactions were generally characterized by deeper minimal depths, suggesting that while these variables are significant in the context of breast cancer risk, their combined influence is considered later in the decision tree model (Fig. 7).

Analysis of LASSO regression model for BC prevalence prediction using WWI and other variables

Based on the LASSO regression results, five variables—age, EA, HTN, MEN, and gravida—were identified as relevant covariates. These variables, together with WWI, were subsequently included in the downstream analyses to examine their associations with BC outcomes (Fig. 8).

Predictive performance and clinical utility of the LASSO model

A nomogram was developed using the final logistic regression model, which incorporated six variables: WWI, MEN, gravida, EA, HTN and age. This tool enables individualized estimation of BC prevalence by summing the points assigned to each variable (Fig. 9).

The calibration curve demonstrated strong agreement between predicted probabilities and observed outcomes, with the bias-corrected line closely following the ideal diagonal. The mean absolute error was 0.001, indicating excellent calibration performance of the model (Fig. 10).

The decision curve analysis showed that the LASSO model incorporating WWI yielded a small net benefit over the"treat-all"and"treat-none"strategies within a threshold probability range of approximately 0.01 to 0.07. Beyond this range, the net benefit of the model diminished, suggesting limited clinical utility under stricter intervention thresholds (Fig. 11).

Comparative evaluation of model performance and calibration

When comparing the performance of the four models (Fig. 12, Table S6), the full model demonstrated the highest discriminative ability, with an AUC of 0.797, sensitivity of 0.865, and specificity of 0.587. The LASSO-based model showed comparable performance (AUC = 0.795, sensitivity = 0.794, specificity = 0.656), followed closely by the random forest model (AUC = 0.790, sensitivity = 0.712, specificity = 0.728), which had an out-of-bag error rate of 3.04%. In contrast, the model incorporating WWI alone yielded a considerably lower AUC of 0.595, with sensitivity and specificity of 0.613 and 0.539, respectively. These findings suggest that WWI, while insufficient as a sole predictor, may contribute complementary value when included in multivariable frameworks.

Multivariable logistic regression using different covariate selection strategies

To evaluate the robustness of the association between WWI and BC, multivariable logistic regression models were constructed using three different covariate selection strategies: (1) manual exclusion based on multicollinearity diagnostics, (2) LASSO regression, and (3) random forest variable importance.

LASSO regression, using tenfold cross-validation, identified five variables with non-zero coefficients: age, educational attainment, hypertension, menopausal status, and gravida (Fig. 8). In parallel, random forest importance rankings were assessed using three methods (Gini index, permutation accuracy, and default measure), and eight variables were consistently ranked among the top: age, gravida, WC, weight, HDL, PIR, BMI, and WWI (Fig. 4). To address potential multicollinearity, generalized variance inflation factor (GVIF) diagnostics were conducted (Table S4, S5). Variables with GVIF > 10 (WC and weight) were excluded from model building. The remaining variables, including WWI, were included in subsequent multivariable logistic regression models constructed based on each selection method. Across all three modeling approaches, WWI was not significantly associated with BC. The detailed model comparisons are presented in Table S7.

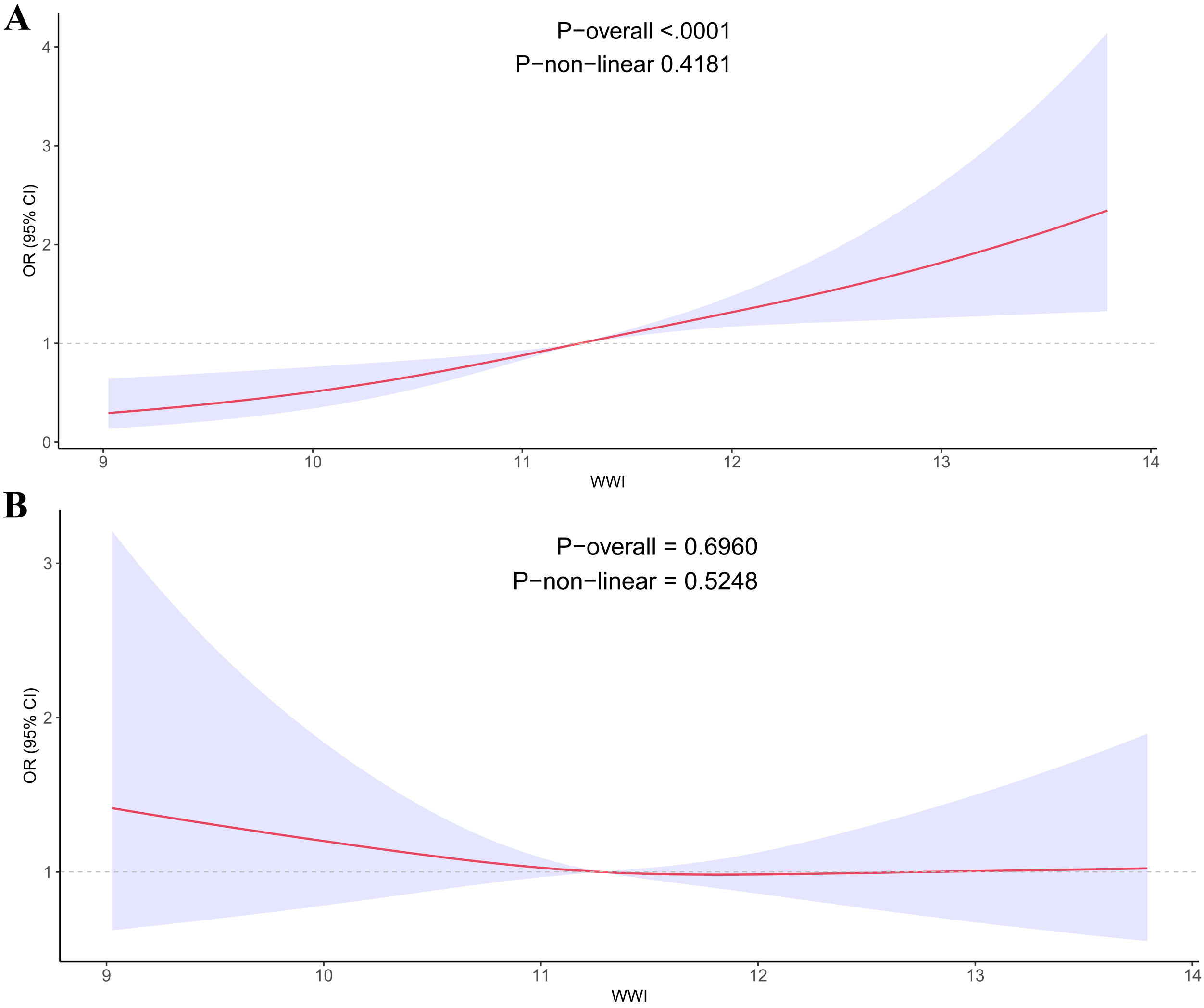
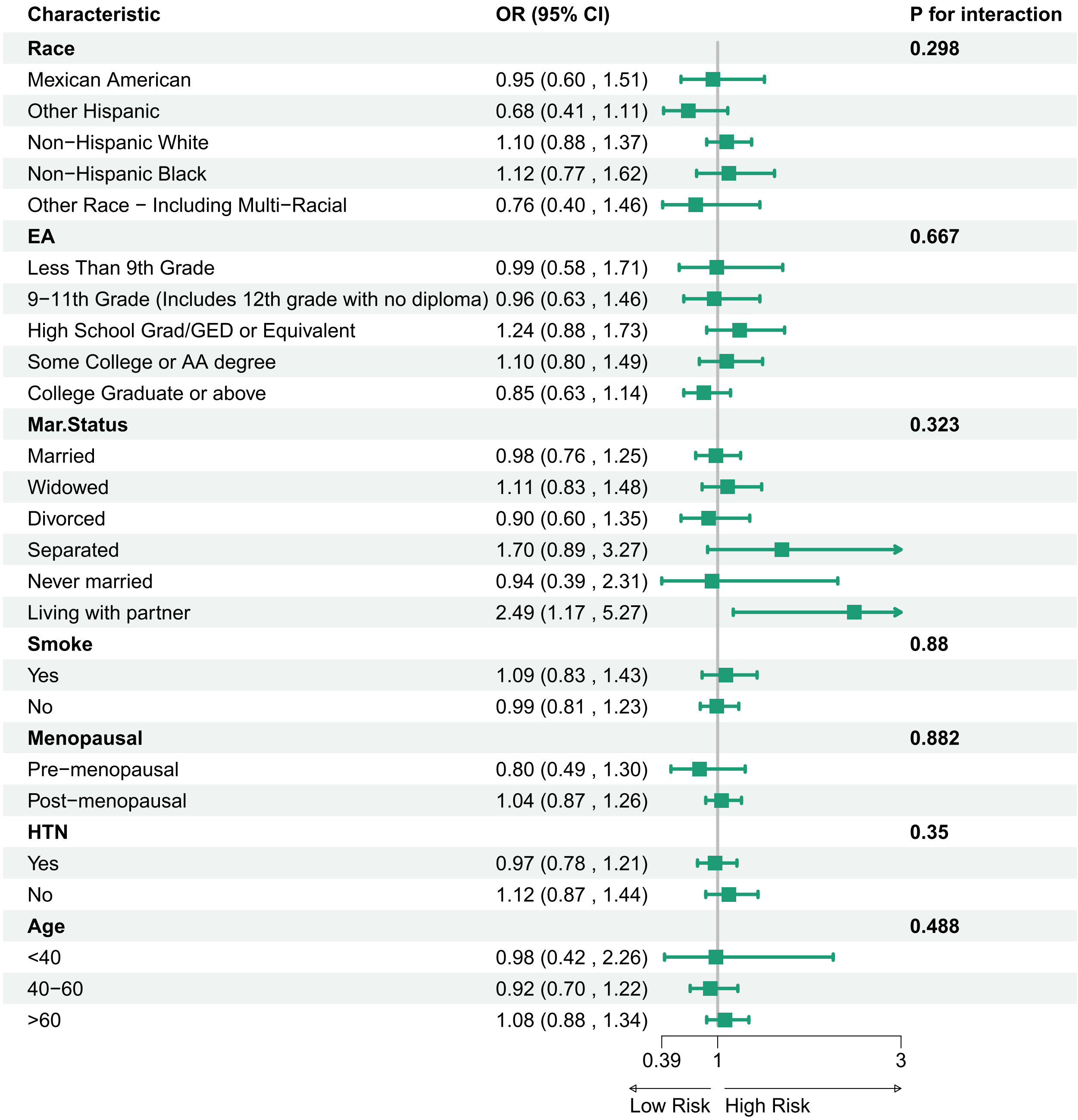


Fig. 2 **A** Restricted cubic splines for the relationship between WWI and BC (unadjusted model). **B** Restricted cubic splines for the relationship between WWI and BC (adjusted model)

Discussion

In this cross-sectional study using data from the 2005–2018 NHANES, we examined the association between the WWI, a recently introduced measure of central adiposity, and the prevalence of BC. Individuals with higher WWI values had increased odds of reporting a BC diagnosis at the time of the survey. In variable importance analyses, WWI ranked higher than BMI based on both the Gini index and default model contribution (Fig. 4), suggesting that WWI may capture relevant information beyond general adiposity in relation to breast cancer prevalence. This supports the potential utility of WWI as a complementary indicator in BC risk profiling.

Our findings differ from those of a previous study that investigated the association between WWI and BC prevalence using NHANES data from 2011 to 2018, which reported a statistically significant relationship [13]. Several methodological differences may account for this discrepancy. First, while Huang et al. analyzed NHANES data from 2011–2018 [13], our study utilized a longer time frame (2005–2018) with a slightly larger sample. Second, the covariates included in the models were not identical. Although both studies focused on women, Huang et al. adjusted for factors such as physical activity, alcohol consumption, coronary heart disease, and diabetes [13], which were not included in our analysis. Our study, on the other hand, included different covariates, such as menopausal status and gravida (number of pregnancies), which may have had an impact on the observed association between WWI and BC. These factors, which were not considered by Huang et al., could play a significant role in BC risk and may have contributed to differences in the findings. Third, we conducted a sensitivity analysis using the covariates specified in Huang et al. to assess the robustness of our findings (Table S8). The result remained non-significant (OR = 0.97; 95% CI:



Characteristic	OR (95% CI)	P for interaction
Race		**0.298**
Mexican American	0.95 (0.60 , 1.51)	
Other Hispanic	0.68 (0.41 , 1.11)	
Non−Hispanic White	1.10 (0.88 , 1.37)	
Non−Hispanic Black	1.12 (0.77 , 1.62)	
Other Race − Including Multi−Racial	0.76 (0.40 , 1.46)	
EA		**0.667**
Less Than 9th Grade	0.99 (0.58 , 1.71)	
9−11th Grade (Includes 12th grade with no diploma)	0.96 (0.63 , 1.46)	
High School Grad/GED or Equivalent	1.24 (0.88 , 1.73)	
Some College or AA degree	1.10 (0.80 , 1.49)	
College Graduate or above	0.85 (0.63 , 1.14)	
Mar.Status		**0.323**
Married	0.98 (0.76 , 1.25)	
Widowed	1.11 (0.83 , 1.48)	
Divorced	0.90 (0.60 , 1.35)	
Separated	1.70 (0.89 , 3.27)	
Never married	0.94 (0.39 , 2.31)	
Living with partner	2.49 (1.17 , 5.27)	
Smoke		**0.88**
Yes	1.09 (0.83 , 1.43)	
No	0.99 (0.81 , 1.23)	
Menopausal		**0.882**
Pre−menopausal	0.80 (0.49 , 1.30)	
Post−menopausal	1.04 (0.87 , 1.26)	
HTN		**0.35**
Yes	0.97 (0.78 , 1.21)	
No	1.12 (0.87 , 1.44)	
Age		**0.488**
<40	0.98 (0.42 , 2.26)	
40−60	0.92 (0.70 , 1.22)	
>60	1.08 (0.88 , 1.34)	

Fig. 3 Subgroup analysis of the association between weight-adjusted waist index (WWI) and breast cancer (BC) risk. Odds ratios (ORs) and 95% confidence intervals (CIs) were obtained from multivariable logistic regression models stratified by subgroup. Each model was adjusted for all covariates listed in the figure—race/ethnicity, educational attainment (EA), marital status (Mar.Status), smoking status, menopausal, hypertension (HTN), and age—except for the stratifying variable itself. All P values for interaction were > 0.05. Abbreviations: EA, educational attainment; Mar.Status, marital status; HTN, hypertension

0.75–1.24; $p = 0.800$), consistent with our main model (OR = 0.98; 95% CI: 0.75–1.26; $p = 0.900$). Finally, while Huang et al. reported a marginally significant association (OR = 1.19; 95% CI: 1.00–1.42; $p = 0.0448$) [13], our RCS analysis found no evidence of a non-linear relationship between WWI and BC prevalence. Our larger sample size, extended study period, and inclusion of additional covariates likely contributed to the attenuation of the observed association.

Some studies have suggested a positive correlation between central obesity and BC risk [14]. Several potential mechanisms may explain this association. Inflammation could be a key mechanism, as fat-induced inflammation and its pro-tumorigenic consequences

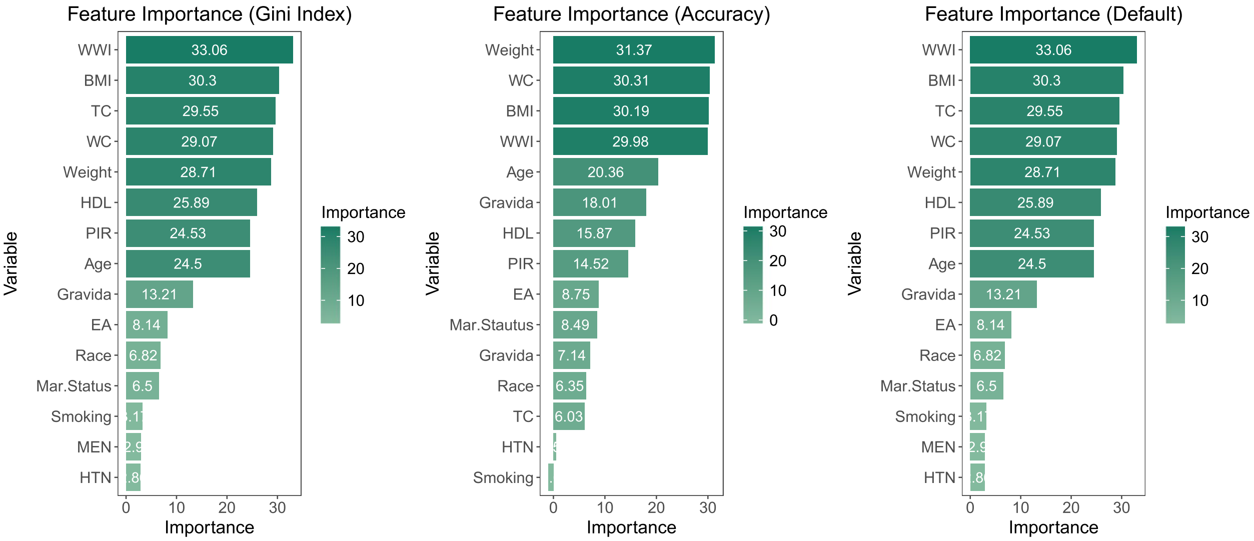


Fig. 4 Variable importance based on Gini index, accuracy loss, and default model contribution. WWI, BMI, and Weight consistently ranked among the top predictors

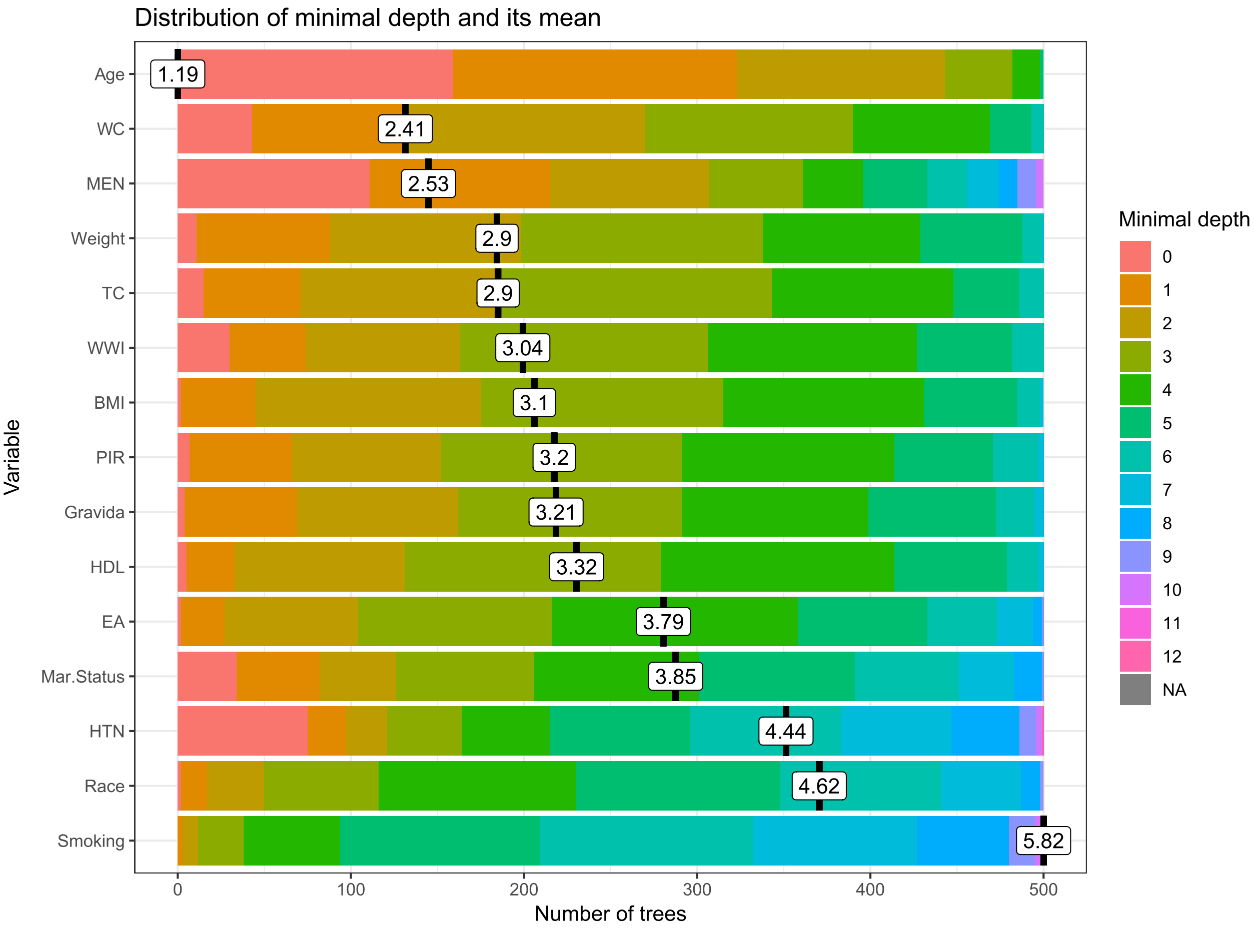


Fig. 5 Minimal depth distribution of variables in the random forest model

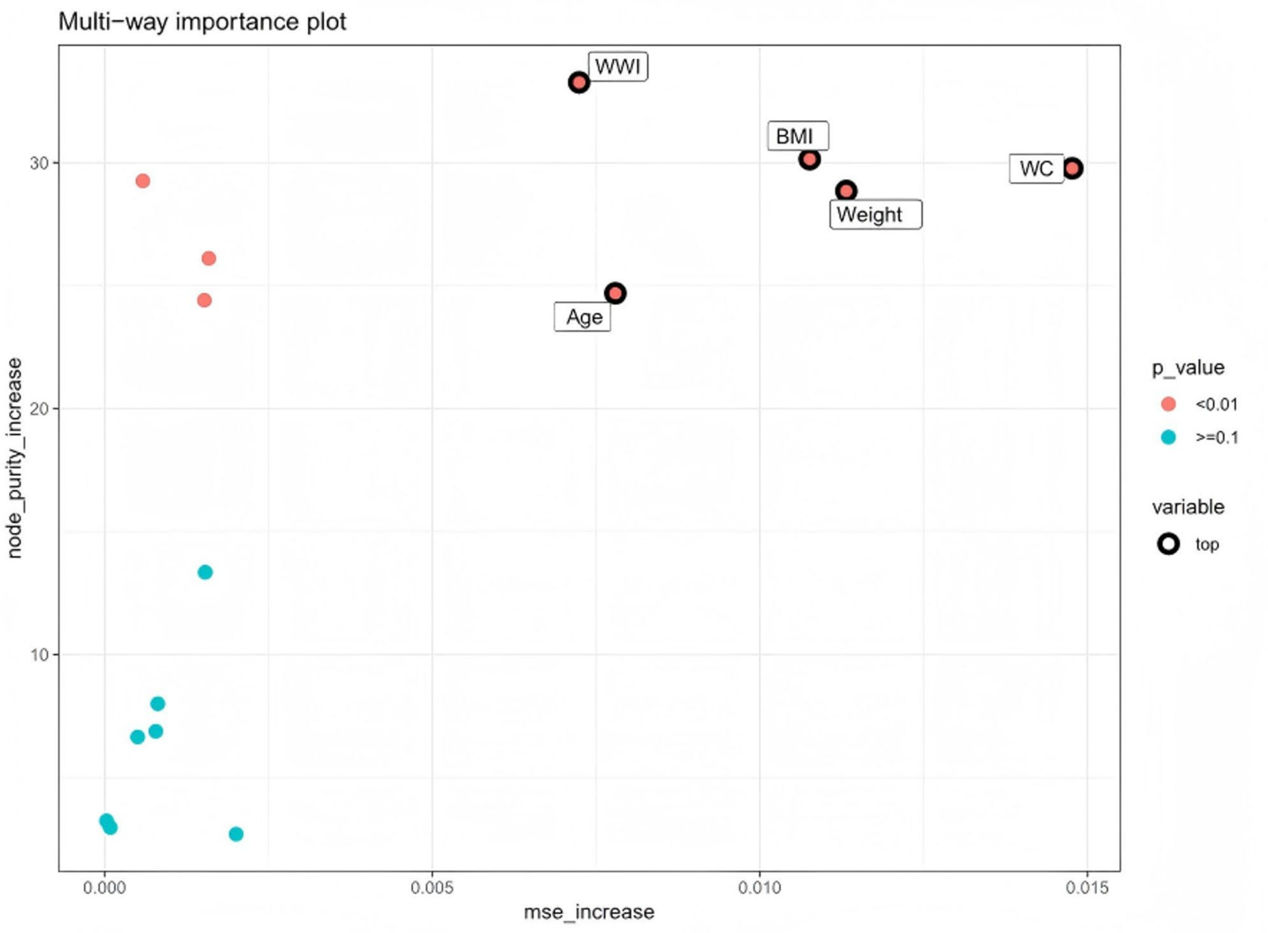


Fig. 6 Multi-way importance plot showing WWI and BMI contributed significantly to node purity and prediction accuracy

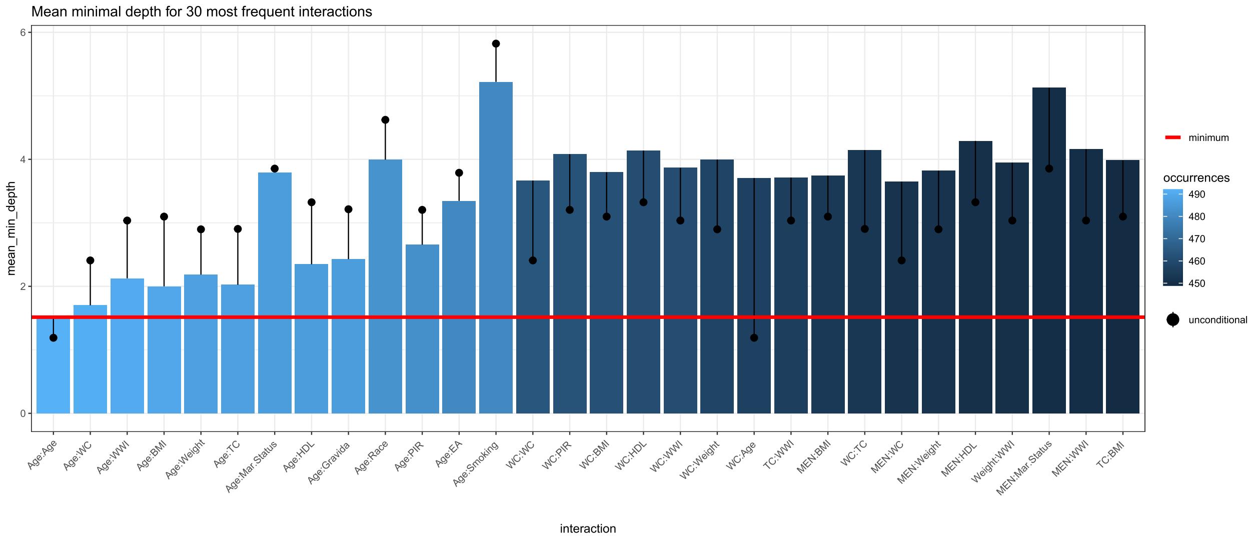


Fig. 7 Variable interaction analysis in the random forest model

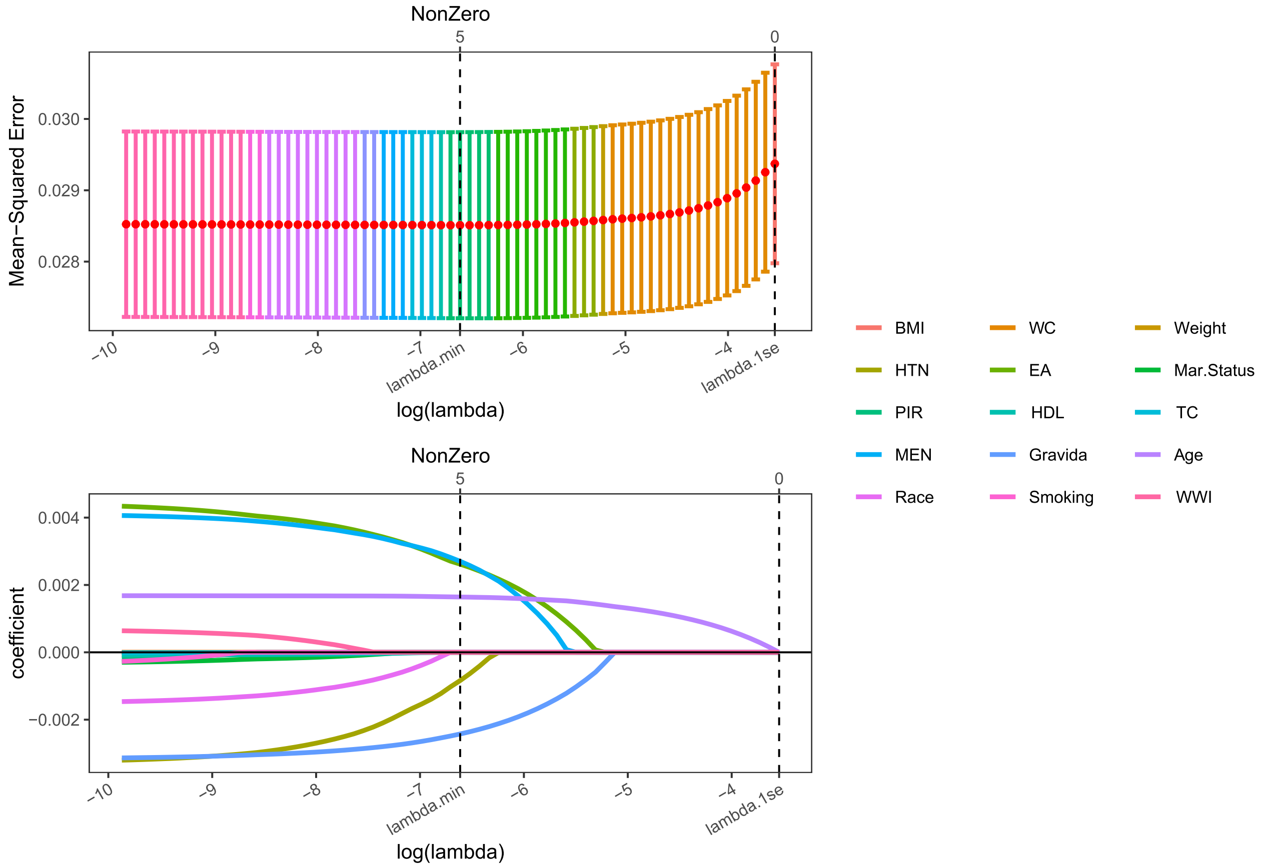


Fig. 8 Variable selection using LASSO regression with ten-fold cross-validation. The upper panel displays the cross-validation curve based on mean squared error (MSE), with the minimum lambda (λ = 0.00134) indicated. The lower panel shows coefficient paths, with five variables retained in the final model

have been observed in individuals who would not be classified as obese or overweight by BMI [5]. Certain inflammatory biomarkers are elevated in individuals with central obesity compared to those without [15], suggesting that central obesity may contribute to increased BC risk by promoting both systemic and local low-grade chronic inflammation, which in turn may enhance pro-tumor oxidative stress [16]. Insulin resistance is another important mechanism. Visceral obesity is strongly associated with hyperinsulinemia and increased levels of insulin-like growth factor 1 (IGF-1), which may stimulate cancer cell proliferation and survival [17]. These metabolic alterations have been linked to both breast cancer incidence and poor prognosis. Central obesity is associated with more invasive tumor characteristics in BC patients [18], which may indicate a poorer prognosis for those with central obesity [19]. Additionally, a study has highlighted that postmenopausal central obesity is linked to a worse prognosis in BC [20].

Central adiposity has gained increasing attention as a critical factor influencing BC incidence, extending beyond traditional obesity measures like BMI. Anthropometric indicators such as WC, waist-hip ratio (WHR), and waist-height ratio (WHtR) have been consistently linked with elevated BC risk in diverse populations [21]. Furthermore, the impact of central obesity on BC risk may vary across different subtypes [22]. Waist circumference is independently positively associated with the risk of Luminal B-HER2 + breast cancer, while hip circumference is independently negatively associated with the risk of this subtype [22]. Notably, several investigations have reported strong associations between increased WC and WHtR and the risk of triple-negative breast cancer (TNBC) [23], a subtype characterized by its aggressive nature. In contrast, these relationships tend to be less pronounced for luminal A tumors, which represent the most common BC subtype and generally have a more favorable prognosis [21]. Additionally, longitudinal evidence indicates that changes in adiposity can influence breast cancer recurrence and survival, highlighting the dynamic role of adipose tissue in disease progression [24].

In our study, we used WWI, a relatively novel metric designed to more accurately reflect central adiposity by adjusting waist circumference for body weight. This is because body weight is a better indicator of overall fat

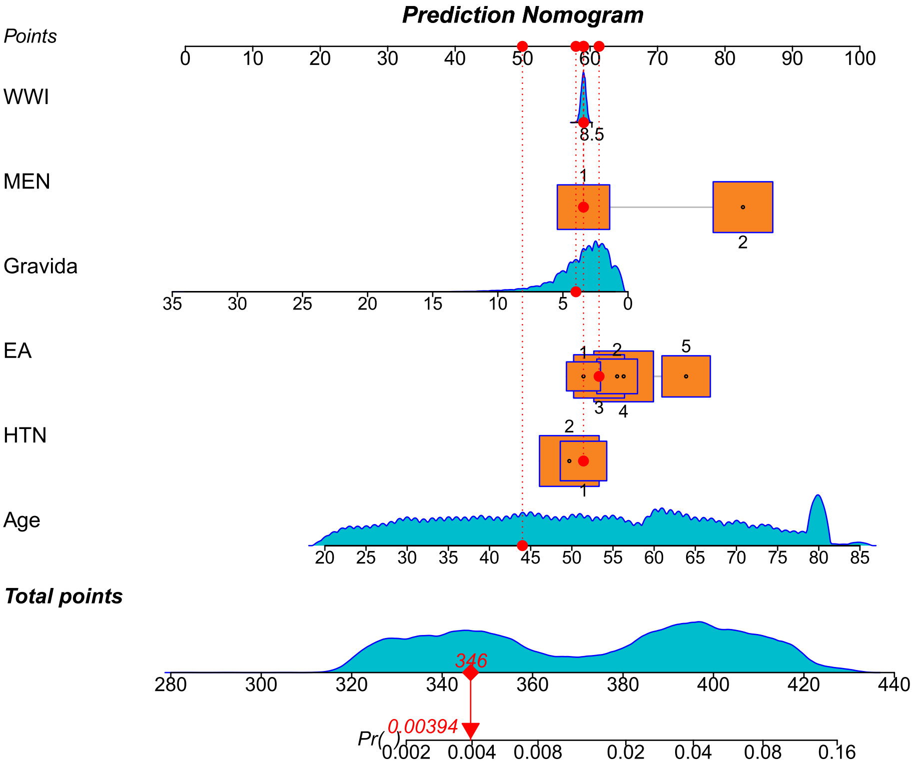


Fig. 9 Nomogram constructed from the LASSO-derived logistic model incorporating WWI, MEN, gravida, EA, HTN and age. The total score corresponds to a predicted breast cancer probability

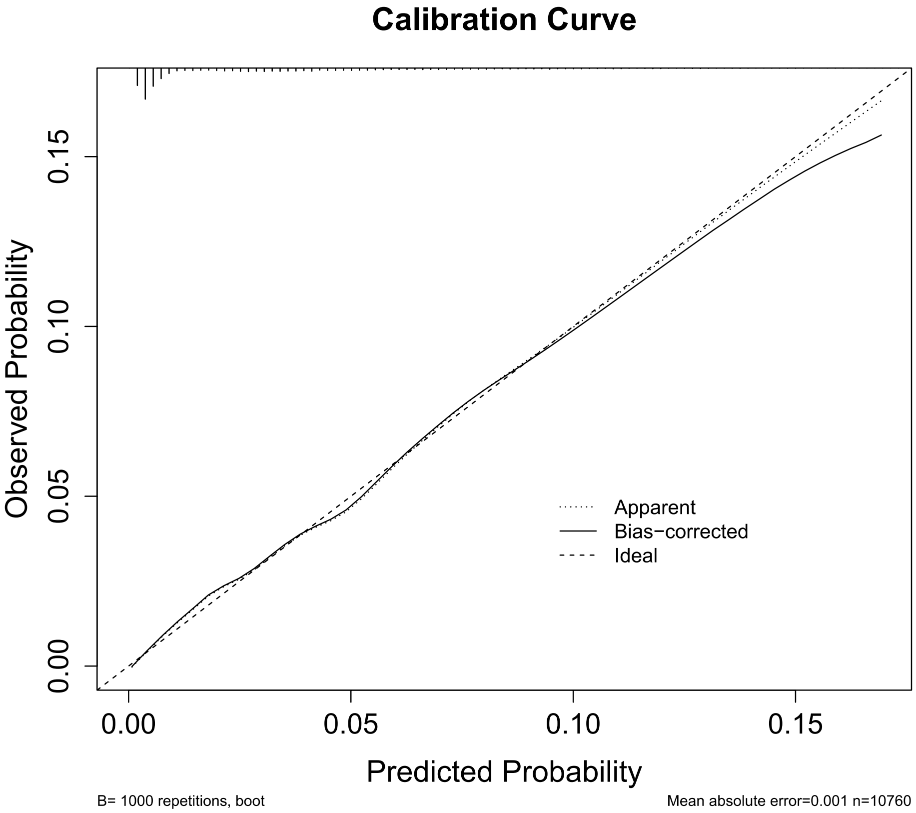


Fig. 10 Calibration curve shows good agreement between predicted and observed probabilities, with minimal deviation after bias correction

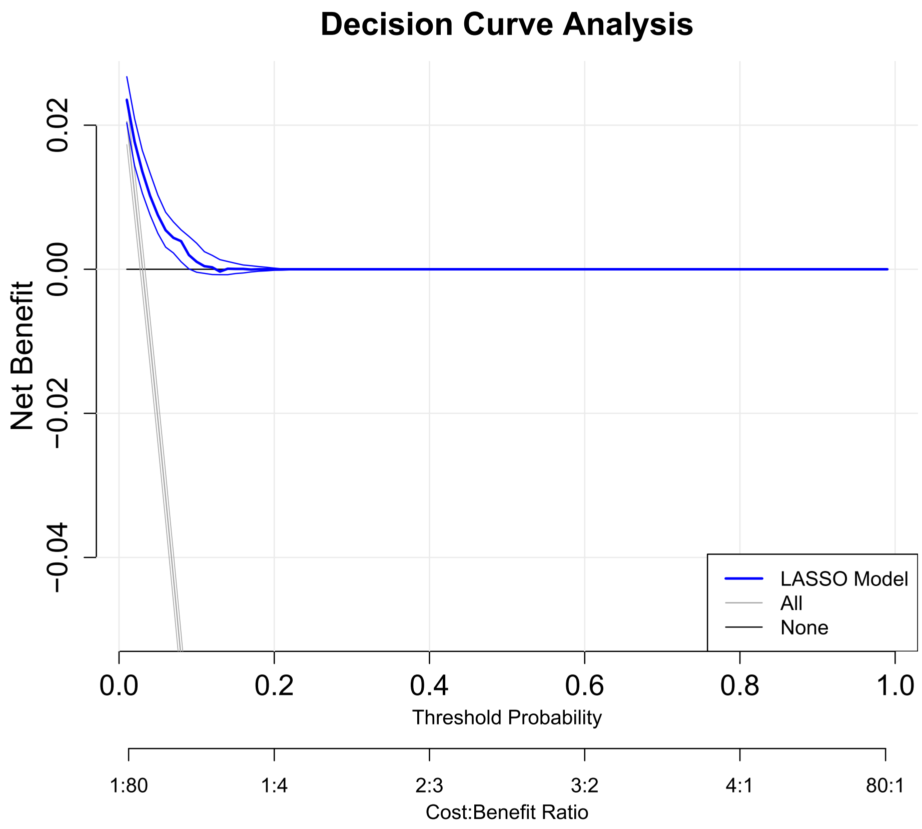


Fig. 11 Decision curve analysis indicates net clinical benefit of using the model across a range of threshold probabilities, especially between 0.02 and 0.1

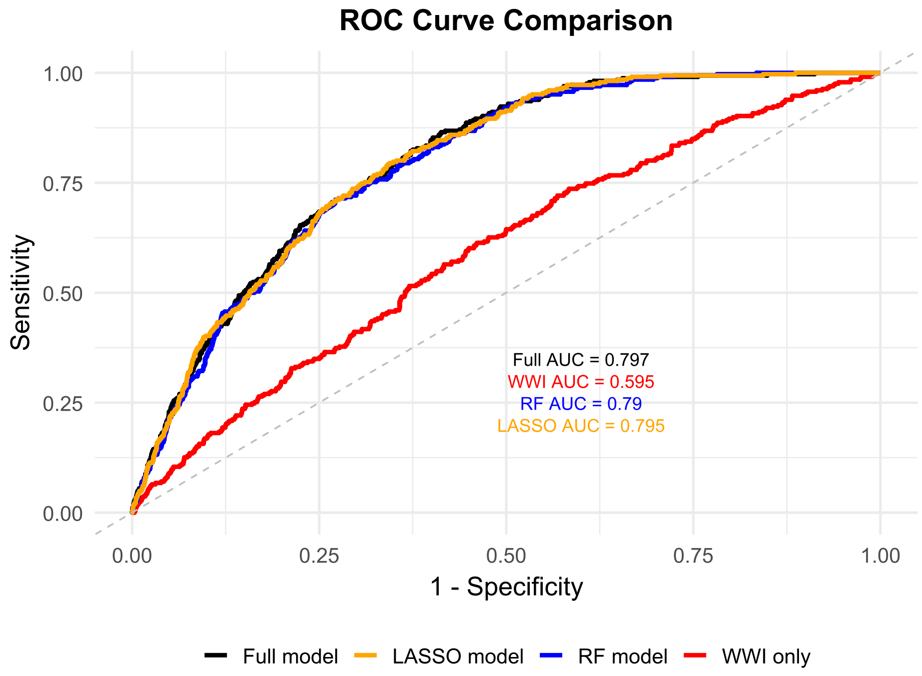


Fig. 12 Receiver operating characteristic (ROC) curves comparing predictive models

mass compared to height, making the waist-to-weight ratio a more reliable measure of central fat distribution. By accounting for body weight, WWI offers a more precise assessment of central adiposity, which is strongly associated with various health risks, including BC.

Collectively, these studies emphasize the significance of central adiposity in BC development and outcomes.

In our study, WWI was initially associated with breast cancer prevalence in both unadjusted and partially adjusted models. However, after full adjustment for relevant confounders, the association was no longer statistically significant. This suggests that WWI may not serve as an independent predictor of BC prevalence. It is important to note that the analytic sample size in this study is relatively small, which may have reduced the power to detect a significant effect in the fully adjusted models. Larger studies with more participants may be required to further investigate the potential role of WWI in BC prediction.

Although this limits WWI's utility as an independent predictor, it may still contribute to broader risk assessment models, particularly in settings where access to comprehensive diagnostics is limited. Its simplicity and correlation with central adiposity may make it a useful component of a multifactorial risk profile, rather than a standalone diagnostic tool.

This study has several limitations. It was based on a U.S. population dataset, in which Asian Americans may be grouped under the "Other" race category, but their sample size is relatively small. Therefore, the findings may have limited generalizability to Asian populations both within and outside the United States due to ethnic diversity and sample representation. Additionally, BC diagnosis in this study was based on self-report rather than registry confirmation, which may introduce information and misclassification bias affecting diagnostic accuracy. However, the inability to examine associations between WWI and specific BC subtypes or stages is not due to the self-reported nature of the data, but rather because NHANES does not collect information on BC molecular subtype or stage. This limitation restricts detailed subtype- or stage-specific analyses. Furthermore, our study analyzes prevalent BC cases rather than incident cases or breast cancer–specific survival. Therefore, interpretations should be cautious, as prevalence-based analyses may be influenced by survival duration and do not establish temporal or causal relationships.

Another potential limitation is the handling of missing data. Although Little's MCAR test indicated that data were not missing completely at random (MCAR), a weighting procedure was applied to reduce bias from non-random missingness. While this adjustment helped mitigate the impact of missing data on generalizability, it is important to acknowledge that residual bias may still exist. Future studies could explore more advanced imputation techniques or alternative approaches to further enhance the robustness and applicability of the findings.

It is also important to note that the cross-sectional design of this study precludes any inference of causality. Because exposures and outcomes were measured simultaneously, we cannot determine the directionality of the observed associations. Prospective longitudinal studies are warranted to validate these findings and better understand the role of WWI in predicting BC incidence and prognosis.

Conclusion

WWI showed a positive association with BC in unadjusted analyses, but this relationship was not maintained after full adjustment for sociodemographic, behavioral, and reproductive factors. While Random Forest machine learning approaches identified WWI as an important variable, its role as an independent predictor of BC prevalence remains unclear.

Given the cross-sectional nature of our study and the relatively small number of breast cancer cases ($N = 326$), these results should be interpreted with caution. The limited sample size may reduce statistical power and affect the generalizability of our findings. Further prospective research with larger sample sizes is needed to clarify the predictive value of WWI and to explore how it might fit into broader models of risk stratification.

Abbreviations

WWI	Weight-adjusted-waist index
BC	Breast cancer
TNBC	Triple-negative breast cancer
NHANES	National Health and Nutrition Examination Survey
OR	Odds ratio
HR	Hazard ratio
CI	Confidence interval
AUC	Area Under the Curve
WC	Waist circumference
BMI	Body mass index
WHR	Waist-hip ratio
WHtR	Waist-height ratio
PIR	Poverty-to-income ratio
TC	Total cholesterol
HDL	High-density lipoprotein
MEN	Menopausal
ROC	Receiver operating characteristic
RCS	Restricted cubic splines
EA	Educational attainment
Mar. Status	Marital status
HTN	Hypertension
LASSO	Least absolute shrinkage and selection operator
GVIFs	Generalized variance inflation factors
MCAR	Missing completely at random

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14651-6.

Supplementary Material 1.

Acknowledgements
The authors would like to acknowledge each of their colleagues.

Authors' contributions
WJW: study design, drafting, plotting, gathering, organizing, analysis and writing. Other authors: interpreting and reviewing. All the authors checked out and approved the final paper.

Funding
Not applicable.

Data availability
The data used in this study are publicly available in the NHANES database (www.cdc.gov/nchs/nhanes). Raw data can be obtained from the corresponding author.

Declarations

Ethics approval and consent to participate
This research employed information gathered from publicly accessible databases; it needed no ethical approval.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]First Affiliated Hospital of Nanchang University, No.17 Yongwai Zhengjie, Nanchang, Jiangxi, China

Received: 24 February 2025 / Accepted: 8 July 2025

References

1. Akram M, Iqbal M, Daniyal M, Khan AU. Awareness and current knowledge of breast cancer. Biol Res. 2017;50:33.
2. Arnold M, et al. Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast Off J Eur Soc Mastology. 2022;66:15–23.
3. Zhou X-D, et al. Burden of disease attributable to high body mass index: an analysis of data from the global burden of disease study 2021. Eclinicalmedicine. 2024;76: 102848.
4. Hu L, et al. Prevalence of overweight, obesity, abdominal obesity and obesity-related risk factors in southern China. PLoS One. 2017;12:e0183934.
5. Iyengar NM, Gucalp A, Dannenberg AJ, Hudis CA. Obesity and cancer mechanisms: tumor microenvironment and inflammation. J Clin Oncol Off J Am Soc Clin Oncol. 2016;34:4270–6.
6. Iyengar NM, et al. Systemic correlates of white adipose tissue inflammation in early-stage breast cancer. Clin Cancer Res. 2016;22:2283–9.
7. Seo BR, et al. Obesity-dependent changes in interstitial ECM mechanics promote breast tumorigenesis. Sci Transl Med. 2015;7(301):301ra130.
8. WHO Expert Consultation. Appropriate body-mass index for asian populations and its implications for policy and intervention strategies. Lancet Lond Engl. 2004;363:157–63.
9. Poirier P. The many paradoxes of our modern world: is there really an obesity paradox or is it only a matter of adiposity assessment? Ann Intern Med. 2015;163:880–1.
10. Flegal KM. Prevalence and trends in obesity among US adults, 1999–2008. JAMA. 2010;303:235.
11. Park Y, Kim NH, Kwon TY, Kim SG. A novel adiposity index as an integrated predictor of cardiometabolic disease morbidity and mortality. Sci Rep. 2018;8:16753.
12. Després J-P. Body fat distribution and risk of cardiovascular disease: an update. Circulation. 2012;126:1301–13.
13. Huang X, et al. Weight-adjusted-waist index: an innovative indicator of breast cancer hazard. BMC Womens Health. 2024;24:660.
14. Houghton SC, Eliassen H, Tamimi RM, Rosner BA, Hankinson SE. Central adiposity and subsequent risk of breast cancer by menopause status. J Natl Cancer Inst. 2021;113:900–8.
15. Chang S-L, Tchernof A, Durocher F, Diorio C. Associations of biomarkers of inflammation and breast cancer in the breast adipose tissue of women with combined measures of adiposity. J Obes. 2021;2021:3620147.
16. Zimta A-A, et al. Molecular links between central obesity and breast cancer. Int J Mol Sci. 2019;20:5364.
17. Zuo Q, Band S, Kesavadas M, Madak Erdogan Z. Obesity and postmenopausal hormone receptor-positive breast cancer: epidemiology and mechanisms. Endocrinology. 2021;162: bqab195.
18. Gioseffi C, de Padilha PC, Chaves GV, de Oliveira LC, Peres WAF. Body weight, central adiposity, and fasting hyperglycemia are associated with tumor characteristics in a brazilian cohort of women with breast cancer. Nutrients. 2022;14:4926.
19. Lubián López DM, et al. Relationship of breast volume, obesity and central obesity with different prognostic factors of breast cancer. Sci Rep. 2021;11:1872.
20. Park Y-MM, et al. The association between metabolic health, obesity phenotype and the risk of breast cancer. Int J Cancer. 2017;140:2657–66.
21. Chen H, et al. The relationship between central obesity and risk of breast cancer: a dose–response meta-analysis of 7,989,315 women. Front Nutr. 2023;10:1236393.
22. Schoemaker MJ, et al. Central and peripheral adiposity and premenopausal breast cancer risk: a pooled analysis of 440,179 women. Breast Cancer Res BCR. 2025;27:55.
23. Chen H-L, Ding A, Wang M-L. Impact of central obesity on prognostic outcome of triple negative breast cancer in Chinese women. Springerplus. 2016;5:594.
24. Pang Y, Wei Y, Kartsonaki C. Associations of adiposity and weight change with recurrence and survival in breast cancer patients: a systematic review and meta-analysis. Breast Cancer Tokyo Jpn. 2022;29:575–88.

Publisher's Note