Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.37326

Clinical and Lifestyle Factors Associated with Breast Cancer History: A Cross-Sectional Analysis Using NHANES Data

Yineng Kuang

University of Toronto, Toronto, Canada
yineng.kuang@mail.utoronto.ca

Abstract. Breast cancer is one of the most common cancers worldwide and remains an important public health concern. Many demographic, clinical, and lifestyle factors have been linked to breast cancer, but these factors are often related to each other, making their independent associations difficult to identify. This study examined factors associated with self-reported breast cancer history using data from the National Health and Nutrition Examination Survey (NHANES). The analysis included 11,933 adults. Demographic characteristics, clinical conditions, and lifestyle factors were first compared between participants with and without a breast cancer history. Multivariable logistic regression was then used to adjust for potential confounding. Older age and smoking history were significantly associated with higher odds of breast cancer history. In contrast, no statistically significant associations were observed for BMI, diabetes, hypertension, alcohol use, physical activity, or sleep duration after adjustment. These results show the importance of considering confounding factors when examining breast cancer risk patterns in population-based data.

Keywords: Breast cancer, NHANES, epidemiology, lifestyle factors, logistic regression

1. Introduction

Breast cancer is one of the most commonly diagnosed cancers worldwide and remains a major cause of cancer-related illness among women [1]. Improvements in screening and treatment have increased survival, but identifying factors related to breast cancer remains important for prevention and early detection [2].

Breast cancer risk is influenced by a combination of demographic, biological, and lifestyle factors. Age is one of the strongest established risk factors, partly because hormonal exposure, genetic changes, and environmental influences accumulate over time [3]. Lifestyle factors also require attention. Smoking exposes individuals to carcinogenic compounds, while obesity may affect estrogen production and metabolic pathways. Alcohol consumption and physical inactivity have also been regarded as potentially modifiable risk factors [4, 5].

However, these factors do not act independently. For example, older adults are also more likely to have hypertension, diabetes, and other chronic conditions. A factor may therefore appear to be associated with breast cancer in a simple comparison even if the relationship is mainly explained by age or another variable. Multivariable analysis is needed to separate these effects.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.37326

This study asks which demographic, clinical, and lifestyle factors are independently associated with self-reported breast cancer history among adults in NHANES. The main predictors examined were age, BMI, diabetes, hypertension, smoking history, alcohol use, physical activity, and sleep duration. A quantitative cross-sectional approach was used, followed by multivariable logistic regression to control for potential confounding.

By examining these variables in the same model, this study aims to provide a comprehensive picture of their independent relationships with breast cancer history and to demonstrate how population-based datasets such as NHANES can be used in cancer epidemiology.

2. Study population and statistical approach

2.1. NHANES dataset and participants

Data were obtained from the National Health and Nutrition Examination Survey (NHANES), a population-based survey conducted in the United States. NHANES collects health information through interviews, physical examinations, and laboratory measurements and is widely used in epidemiological research [4, 5].

A total of 11,933 adult participants were included in this analysis. The primary outcome was self-reported breast cancer history. Participants who reported a previous breast cancer diagnosis were classified as having a breast cancer history, while all others formed the comparison group.

The predictors were grouped into three general areas. Demographic factors included age. Clinical factors included BMI, diabetes, and hypertension. Lifestyle factors included smoking history, alcohol consumption, physical activity, and sleep duration. These variables were selected because previous research has suggested possible relationships with breast cancer or related health outcomes [3-5].

2.2. Statistical analysis

The analysis was conducted in two stages. First, descriptive statistics were used to summarize participant characteristics. Continuous variables were compared between groups using appropriate statistical tests, while categorical variables were assessed using chi-square tests.

These initial comparisons were useful for identifying possible associations, but they could not account for confounding. Age is a clear example. Older participants are more likely to develop breast cancer, but they are also more likely to have diabetes and hypertension. Therefore, an observed relationship between these conditions and breast cancer history could partly reflect age-related differences.

To address this issue, multivariable logistic regression was used because the outcome variable was binary. The model included demographic, clinical, and lifestyle variables simultaneously. Results were expressed as odds ratios (ORs), which represent the association between each predictor and breast cancer history after adjustment for the other variables. Table 1 summarizes the weighted characteristics of the study population, stratified by breast cancer history status, with comparisons made using survey-weighted t-tests and chi-square tests. These analyses were intended to describe the sample characteristics and identify potential associations; however, because they did not account for potential confounding, they should be considered preliminary, and the main conclusions of this study are based on the multivariable-adjusted results reported in Chapter 4.

DOI: 10.54254/2753-8818/2026.37326

Table 1. Weighted characteristics of the study population by breast cancer history

Variable	level	0	1	p	test
n		11730	203		
RIDAGEYR (mean, SD)		37.78 (25.46)	69.37 (9.45)	<0.001	
RIAGENDR	1	5575 (47.5)	0 (0.0)	<0.001	
	2	6155 (52.5)	203 (100.0)		
RIDRETH3	1	1114 (9.5)	3 (1.5)	<0.001	
	2	1360 (11.6)	13 (6.4)		
	3	6078 (51.8)	139 (68.5)		
	4	1575 (13.4)	22 (10.8)		
	6	672 (5.7)	9 (4.4)		
	7	931 (7.9)	17 (8.4)		
DMDEDUC2	1	364 (4.8)	9 (4.4)	0.929	
	2	645 (8.5)	21 (10.3)		
	3	1705 (22.5)	44 (21.7)		
	4	2311 (30.4)	59 (29.1)		
	5	2555 (33.7)	70 (34.5)		
	9	11 (0.1)	0 (0.0)		
INDFMPIR (mean, SD)		2.70 (1.67)	3.02 (1.63)	0.015	
BMXBMI (mean, SD)		27.19 (8.13)	30.53 (7.80)	<0.001	
diabetes	0	10230 (90.8)	141 (74.6)	<0.001	
	1	1033 (9.2)	48 (25.4)		
hypertension	0	5435 (65.6)	83 (40.9)	<0.001	
	1	2849 (34.4)	120 (59.1)		
smoking	0	4774 (60.3)	104 (51.2)	0.011	
	1	3144 (39.7)	99 (48.8)		
alcohol	0	547 (10.2)	11 (8.2)	0.533	
	1	4795 (89.8)	123 (91.8)		
physical_activity	0	1532 (19.4)	63 (31.0)	<0.001	
	1	6349 (80.6)	140 (69.0)		
SLD012 (mean, SD)		7.75 (1.61)	8.07 (1.65)	0.005	

DOI: 10.54254/2753-8818/2026.37326

3. Demographic and clinical factors

3.1. Age

Age showed the strongest association with breast cancer history. Participants who reported a previous breast cancer diagnosis were generally older than participants without such a history as shown in Table 2. This pattern is consistent with existing epidemiological evidence showing that breast cancer incidence increases with age [2].

Several biological processes may contribute to this relationship. Over time, individuals experience longer exposure to hormonal and environmental influences and accumulate more genetic changes. Cellular repair mechanisms may also become less efficient with aging [3].

Importantly, age remained significant in the multivariable model. Each additional year of age was associated with approximately an 8.8% increase in the odds of having a breast cancer history. Because the association remained after adjustment for the other variables, age appears to be an independent factor in this analysis.

Table 2. Comparison of age between participants with and without a history of breast cancer using Welch's two-sample t-test

Variable	No breast cancer Mean	Breast cancer Mean	Mean difference	95% CI	t value	df	p value
Age (years)	37.78	69.37	31.59	-32.98 to -30.21	-44.91	255.92	<0.001

3.2. BMI, diabetes, and hypertension

BMI, diabetes, and hypertension were also examined. Obesity has previously been linked to breast cancer through several possible mechanisms, including estrogen production in adipose tissue, insulin resistance, and chronic inflammation. The relationship may be particularly relevant among postmenopausal women [4].

In the current analysis, however, BMI was not statistically significant after multivariable adjustment. This does not necessarily conflict with previous research. The relationship between body weight and breast cancer may depend on age, menopausal status, hormonal factors, and other characteristics that were not fully captured in this analysis.

Diabetes and hypertension were also not statistically significant in the adjusted model. Both conditions are more common at older ages and often occur together with obesity and other metabolic risk factors. Their apparent associations with breast cancer in simpler comparisons may therefore be partly explained by these shared characteristics.

Another issue is that NHANES is cross-sectional. BMI and current health conditions were measured at the time of the survey, but breast cancer may have been diagnosed years earlier. As a result, the analysis cannot determine whether these clinical characteristics predated the cancer diagnosis.

DOI: 10.54254/2753-8818/2026.37326

4. Lifestyle factors

4.1. Smoking history

Among the lifestyle variables, smoking history remained significantly associated with breast cancer history after adjustment. Participants with a smoking history had approximately 55% higher odds of having a history of breast cancer than those without a smoking history as shown in Table 3.

This association is biologically plausible because tobacco smoke contains carcinogenic compounds that can cause DNA damage and oxidative stress. Previous studies have also reported increased breast cancer risk among some groups of active or long-term smokers [5].

The result suggests that smoking history may have an association with breast cancer that is not fully explained by age or the other variables included in the model. However, the NHANES measure used here is based on self-reported smoking history. More detailed information on smoking duration, intensity, and age at initiation would allow the relationship to be examined more precisely.

Table 3. Comparison of smoking history between participants with and without a history of breast cancer

Variable	No breast cancer n (%)	Breast cancer n (%)	Chi-square	df	p value
Smoking			6.4	1	0.011
- No	4774 (60.3%)	104 (51.2%)			

4.2. Alcohol, physical activity, and sleep

Alcohol consumption, physical activity, and sleep duration were not statistically significant in the adjusted regression model. This was somewhat different from what might be expected based on previous research. Alcohol has been associated with breast cancer through hormonal and other biological pathways, while physical activity may reduce risk through effects on body weight and metabolism [4].

One possible explanation is the way these behaviors were measured. Lifestyle patterns can change considerably over a person's lifetime. Current alcohol intake or physical activity may therefore differ from behaviors before breast cancer developed.

Sleep duration has a similar limitation. A measurement collected during one survey period may not reflect long-term sleep patterns. For these variables, longitudinal data collected before cancer diagnosis would provide stronger evidence than a cross-sectional measurement.

5. Multivariable findings and interpretation

The multivariable model produced a clearer picture than the initial comparisons. After age, clinical characteristics, and lifestyle factors were considered together, only age and smoking history remained statistically significant predictors of breast cancer history as shown in Figure 1 and Table 4.

This difference between unadjusted comparisons and the adjusted model illustrates why confounding is important. Diabetes or hypertension, for example, may appear to be related to breast cancer because all three conditions become more common with increasing age. Once age is included in the regression model, the apparent association of these conditions may become much weaker.

DOI: 10.54254/2753-8818/2026.37326

At the same time, a non-significant result should not be interpreted as proof that a factor has no relationship with breast cancer. Cancer development is complex, and the effects of BMI, alcohol, physical activity, or metabolic conditions may vary across different groups. Some associations may also be difficult to detect because NHANES provides a cross-sectional snapshot rather than long-term exposure data.

The results are therefore most useful for identifying patterns within this NHANES sample rather than establishing causal relationships. They also show the advantage of analyzing several risk domains together instead of considering each variable separately.

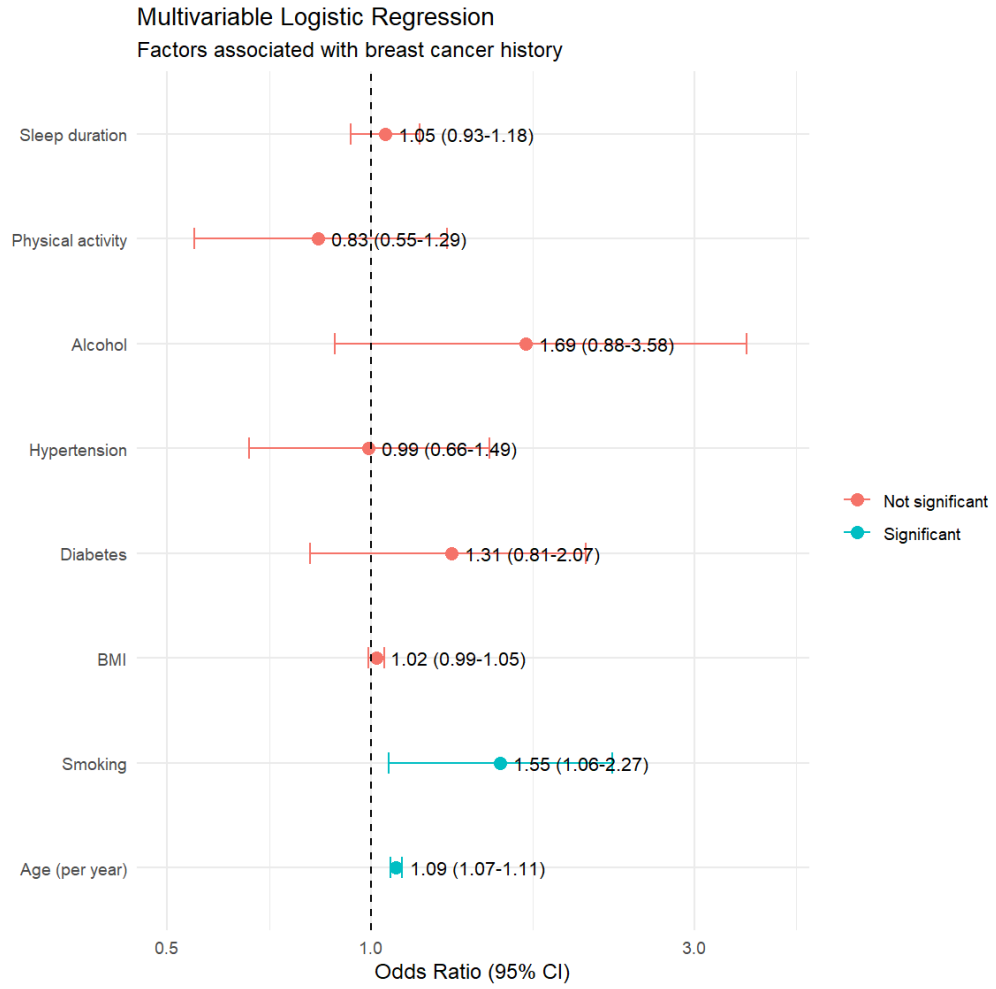

Figure 1. Forest plot of multivariable logistic regression analysis for factors associated with breast cancer history

Table 4. Multivariable logistic regression analysis of factors associated with breast cancer history

Variable	OR	95% CI	P-value
Age	1.09	1.07 - 1.11	<0.001 ***
Sex	100751315.99	0.00 - Inf	0.973
BMI	1.02	0.99 - 1.05	0.153
Diabetes	1.31	0.82 - 2.09	0.253

DOI: 10.54254/2753-8818/2026.37326

Table 4. (continued)

Hypertension	0.99	0.66 - 1.49	0.967
Smoking	1.55	1.06 - 2.27	0.024 *
Alcohol consumption	1.69	0.85 - 3.38	0.138
Physical activity	0.83	0.54 - 1.28	0.409
Sleep duration	1.05	0.93 - 1.18	0.421

6. Conclusion

This study examined demographic, clinical, and lifestyle factors associated with self-reported breast cancer history among 11,933 adults in NHANES. Multivariable logistic regression was used to determine whether observed associations remained after adjustment for potential confounding.

Age and smoking history were the two factors that remained statistically significant. Each additional year of age was associated with approximately 8.8% higher odds of breast cancer history, while participants with a smoking history had about 55% higher odds than those without a smoking history. BMI, diabetes, hypertension, alcohol use, physical activity, and sleep duration were not statistically significant after multivariable adjustment.

These findings should be interpreted with several limitations in mind. Most importantly, the cross-sectional design cannot establish whether an exposure occurred before or after breast cancer diagnosis and therefore cannot demonstrate causality. Breast cancer history and several lifestyle variables were also self-reported, which may introduce recall or reporting bias. In addition, information on tumor subtype, stage, treatment, genetic factors, and detailed reproductive history was limited or unavailable.

Future studies could use longitudinal data to examine exposures before breast cancer diagnosis and include more detailed information on genetics, reproductive history, medication use, and long-term lifestyle patterns. Within the limits of the current dataset, the results show that age and smoking history have the clearest independent associations with breast cancer history and highlight the importance of controlling for confounding when studying cancer risk factors.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249.

[2] Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians, 73*(1), 17–48.

[3] Colditz, G. A., Bohlke, K., & Berkey, C. S. (2014). Breast cancer risk accumulation starts early: Prevention must also. *Breast Cancer Research and Treatment, 145*(3), 567–579.

[4] World Health Organization. (2024). Breast cancer. *World Health Organization*.

[5] Gaudet, M. M., Carter, B. D., Brinton, L. A., Falk, R. T., Gram, I. T., Luo, J., & Milne, R. L. (2018). Pooled analysis of active smoking and breast cancer risk in the Cancer and Steroid Hormone Study and the Nurses' Health Study. *Cancer Prevention Research, 11*(8), 489–500.