Comprehensive Journal of Science
Volume (9), Issue (36), (Sept 2025)
ISSN: 3014-6266

مجلة العلوم الشاملة
المجلد(9) العدد (36) (سبتمبر 2025)
ردمد: 3014-6266

Risk Factors Associated with Breast Cancer Stage at Diagnosis in Libya: A Multivariate Statistical Application and Analysis

Mohamed Albahloul Musrati,
Department of Sstatistics, Faculty of Science, University of Tripoli,

Received: 2-08-2025; Revised: 08-09-2025; Accepted: 13-09-2025; Published 10 -09-2025

Abstract

Background: Breast cancer is the most common malignancy among women in Libya and is frequently diagnosed an advanced stage, adversely affecting prognosis and survival. Evidence on factors independently associated with late-stage presentation in Libya remains limited, particularly from studies employing multivariate statistical methods.

Objectives: This study aimed to identify the demographic, reproductive, behavioral, and biological factors independently associated with advanced-stage breast cancer diagnosis (Stage III) among Libyan women.

Methods: An analytical cross-sectional study was conducted among 220 women with histologically confirmed breast cancer treated at the Al-Riaya Medical Clinic, Tripoli. Data were obtained from structured interviews, medical records, and laboratory databases covering the period 2017–2022. The outcome variable was stage at diagnosis, dichotomized into early stage (Stages I–II) versus advanced stage (Stage III). Multivariate logistic regression analysis was performed to estimate adjusted odds ratios (AORs) and 95% confidence intervals (CIs) for factors associated with Stage III diagnosis.

Results: Of the 220 patients, 44% were diagnosed at Stage III. After adjustment for potential confounders, nulliparity was the strongest predictor of advanced-stage diagnosis (AOR = 6.42, 95% CI: 2.87–14.36; $p < 0.001$). Symptom-based detection was associated with nearly a fivefold increased likelihood of Stage III disease compared with screening-based diagnosis (AOR = 4.91, 95% CI: 2.24–10.77; $p < 0.001$). Physical inactivity (AOR = 3.15, 95% CI: 1.45–6.84; $p = 0.004$), positive family history of cancer (AOR = 2.78, 95% CI: 1.33–5.82; $p = 0.007$), and vitamin D deficiency (AOR = 2.21, 95% CI: 1.12–4.35; $p = 0.022$) were also independently associated with advanced-stage diagnosis. Elevated fasting blood glucose showed a positive but non-significant association with Stage III disease.

Keywords: Breast cancer; Stage at diagnosis; Risk factors; Logistic regression; Libya; Multivariate analysis

1. Introduction

Breast cancer is the most frequently diagnosed malignancy among women worldwide, accounting for approximately 25% of all newly diagnosed female cancer cases [1]. Despite advances in screening, diagnosis, and treatment, breast cancer remains a major public health challenge, particularly in low-

and middle-income countries where late-stage diagnosis is common and survival outcomes are poorer.

In Libya, breast cancer represents the leading cancer among women. According to national cancer registry data, nearly 6,000 new cases are diagnosed annually, constituting approximately 23.7% of all female cancers [2]. The age-standardized incidence rate in Libya has been reported at about 18.8 per 100,000 females, a rate comparable to some African countries such as Nigeria, but considerably lower than rates observed in many European countries, including Finland. Despite this relatively lower incidence, breast cancer in Libya is characterized by a high proportion of diagnoses at advanced stages (predominantly stages II and III), which significantly worsens prognosis, limits treatment options, and increases healthcare costs [3].

Notably, breast cancer in Libya tends to occur at younger ages compared to Western populations. The average age at diagnosis is approximately 45–47 years, reflecting a predominance of premenopausal cases, whereas in Europe the majority of cases occur between 45 and 69 years. Furthermore, Libyan patients often present with more aggressive disease characteristics, including larger tumor size, higher histological grade, and frequent lymph node involvement. These features are associated with poorer clinical outcomes and reduced survival compared to patients in high-income settings. Survival rates in Libya have been reported to lie between those of Nigeria, which has relatively low survival, and Finland, which has among the highest survival rates globally [4].

Within the broader Middle East and North Africa (MENA) region, breast cancer is also the most common cancer among women. In 2019, an estimated 835,576 new cases and 35,405 breast cancer–related deaths were reported in North Africa and the Middle East. Although incidence rates in the MENA region are generally lower than those in Western countries, mortality rates are often comparable or even higher, largely due to delayed diagnosis and limited access to early detection and optimal treatment. Women in the region are typically diagnosed nearly a decade earlier than their Western counterparts. Behavioral and metabolic risk factors—such as obesity, low physical activity, diets high in fat and red meat, and elevated fasting plasma glucose—have been shown to significantly influence breast cancer incidence and mortality in the region [5][6].

From a global perspective, Africa reported an age-standardized breast cancer incidence rate of approximately 37.9 per 100,000 females in 2018, which is higher than that reported for Libya. The average mortality-to-incidence ratio in Africa is about 0.44, reflecting poor survival in many countries. Libya, by contrast, has a lower mortality-to-incidence ratio (approximately 0.24), suggesting relatively better survival compared to several sub-Saharan African countries. Nevertheless, like other developing and transitioning regions, Libya continues to experience rising breast cancer incidence alongside persistently elevated mortality, largely attributable to late-stage diagnosis, limited screening coverage, and health system constraints [7][8].

Stage at diagnosis remains one of the strongest predictors of breast cancer survival. Numerous studies have identified factors associated with advanced-stage presentation, including demographic characteristics (such as age and socioeconomic status), clinical features (tumor grade, histological type, hormone receptor status), behavioral factors (breast self-examination and participation in screening programs), and biological markers (including vitamin D deficiency and elevated blood glucose levels) [9][10]. However, most existing evidence is derived from high-resource settings, and findings may not be directly transferable to the Libyan context due to differences in population structure, health-seeking behavior, and healthcare systems.

From a methodological standpoint, investigating determinants of stage at diagnosis poses significant statistical challenges because these factors are multidimensional and interrelated. Reliance on univariate or bivariate analyses may obscure true associations and lead to biased or misleading conclusions. Multivariate statistical approaches - particularly logistic regression models - are therefore essential to control for confounding variables and to identify independent predictors of late-stage breast cancer diagnosis [11].

Despite a limited number of studies describing the epidemiology and clinical features of breast cancer in Libya [2][12], there is a clear lack of research applying advanced multivariate statistical techniques to examine factors independently associated with late-stage diagnosis. To the best of the authors' knowledge, no previous Libyan study has comprehensively addressed this issue using a multivariate analytical framework.

Accordingly, this study aims to investigate the risk factors associated with the stage of breast cancer at diagnosis in Libya. The specific objectives are to:

1. Describe the demographic, clinical, and biological characteristics of breast cancer patients in Libya.
2. Identify factors independently associated with advanced-stage breast cancer diagnosis (Stage III) using multivariate statistical models.
3. Quantify the strength and direction of these associations by estimating odds ratios with corresponding confidence intervals.

The remainder of this paper is organized as follows: Section 2 describes the materials and methods; Section 3 presents the results; Section 4 discusses the findings; and Section 5 concludes with recommendations and policy implications.

2. Materials and Methods

2.1 Study Design and Setting

An analytical cross-sectional study was conducted at the Oncology Department of Al-Riaya Medical Clinic in Tripoli, Libya. This clinic is one of the main oncology treatment centers in the region, making the study sample highly relevant to the Libyan healthcare context.

2.1.1 Study Population and Sample

The study population comprised all women with histologically confirmed breast cancer who presented for treatment at Al-Riaya Medical Clinic during the specified study periods. The final analytical sample included a total of 220 patients. Of these, 100 patients were enrolled through consecutive sampling between January 2021 and December 2022 and constituted the comprehensive data group; for these patients, information was collected using structured interviews with a predesigned questionnaire in addition to a detailed review of medical records. To increase the statistical power of analyses involving laboratory biomarkers, an additional 120 patients were selected using simple random sampling from the clinic's laboratory database covering the period 2017–2022. This laboratory data group provided information on serum cancer antigen CA 15-3, vitamin D levels, and fasting blood glucose.

Eligibility criteria were defined a priori. Women were included in the study if they had a histologically confirmed diagnosis of breast cancer of any histological type, were of female sex, and had sufficient clinical information available to determine the stage at initial diagnosis, classified according to the TNM system or a simplified clinical staging scheme (Stages I–III). Patients were excluded if they presented with distant metastases at diagnosis (Stage IV), had recurrent breast cancer, or had medical records with incomplete or missing essential data required for analysis.

2.1.2 Study Variables

Data were obtained from structured patient interviews, medical records, and laboratory databases. The study variables were categorized as follows:

Demographic and Socioeconomic Variables include age (years), sex, body weight (kg), height (cm), occupation, self-reported economic status (classified as low, middle, or high), marital status, age at marriage, and age at first childbirth.

Reproductive Health information was collected on age at menarche, age at menopause, regularity of the menstrual cycle, parity (number of children), age at first childbirth, type of breastfeeding (exclusive or mixed), and history of hormonal contraceptive use. Behavioral and familial risk factors included family history of cancer among first- and second-degree relatives, level of physical activity (low, moderate, or high), and dietary habits, specifically the consumption of fruits and vegetables, saturated meats, and fast food. Diagnostic information comprised the reason for breast cancer detection (routine screening versus symptom-based presentation) and the stage of breast cancer at initial diagnosis, classified into stage I, stage II, or stage III. Laboratory data included serum levels

of cancer antigen CA 15-3 measured before and after surgery (U/mL), serum vitamin D concentration (ng/mL), and fasting blood glucose levels (mg/dL).

2.2 Statistical Approaches for Analyzing Risk Factors

logistic regression models allow medical researchers to help clinicians in the choice of an appropriate treatment strategy for individual patients. Multiple logistic regression is a common multivariate statistical technique for modeling the association between a categorical dependent variable and several independent variables. In this context, logistic regression calculates the likelihood of a binary result based on several explanatory variables, rendering it crucial for empirical research involving dichotomous outcomes, such as the presence or absence of a disease or the success or failure of a treatment.

Numerous logistic regressions have gained significant popularity in medical and epidemiological research due to its capacity to manage numerous risk factors concurrently while delivering interpretable outputs, including odds ratios and their corresponding confidence intervals. It represents the log-odds of the event as a linear amalgamation of predictors, incorporating continuous, categorical, and interaction factors [13][14].

We performed a logistic regression models to identify factors independently associated with advanced-stage (Stage III) breast cancer diagnosis. Adjusted odds ratios (AORs) and 95% confidence intervals (CIs) were estimated.

Specification of Model Variable

Outcome Variable (Binary Response)

Let

$$W_i = \begin{cases} 1, & \text{if patient i is diagnosed at advanced stage (Stage III)} \\ 0, & \text{if patient i is diagnosed at early stage(Stage I– II} \end{cases}$$

Let

$$X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$$

denote a vector of k explanatory variables. The logistic regression model is expressed as:

$$P_r(W_i = 1|X_i) = p_i$$

with

$$\text{Log} \quad \frac{p_i}{(1-p)_i} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

Equivalently, the probability of being diagnosed at stage III is:

$$p_i = \frac{\exp\left(\beta_0 + \sum_{j-1}^{k} \beta_j X_{ij}\right)}{1 + \exp\left(\beta_0 + \sum_{j-1}^{k} \beta_j X_{ij}\right)}$$

Where p_i represents the probability of being diagnosed with advanced-stage breast cancer (stage III**),** and $\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right)$ is th odd of breast cancer. The parameter β_0 is the intercept of the model intercept (log-odds of Stage III when all covariates are zero), β_j is the log-odds change in advanced-stage diagnosis associated with a one-unit increase in X_i, adjusted for all other variables and $\exp(\beta_j)$ is the adjusted odds ratio (AOR) for predictor X_j .

The hypothesis testing for independent association is ass follows: for each predictor X_j:

$$H_{0:}\ \beta_j = 0$$

$$H_{1:}\ \beta_j \neq 0$$

Significance is typically assessed using Wald test and Likelihood ratio test.

The logistic regression model estimates the effect of each risk factor on the likelihood of breast cancer by quantifying associations in terms of odds ratios (ORs) with corresponding 95% confidence intervals (CIs)**.** An odds ratio greater than one indicates an increased likelihood of the outcome, whereas an odds ratio less than one indicates a protective effect.

3. Results and discussion

Socio-demographic and Clinical Characteristics of the Study Population

Among the 220 patients included in the study, the mean age was 54.3 ± 12.7 years, with an age range of 25 to 95 years. The age distribution showed that 30% of patients were in the 45–55-year age group, followed by 26% in the 55–65-year group. The study population was predominantly female (97%).

Regarding socio-economic characteristics, 55% of the participants were classified as having a middle economic status, and 49% were married. A positive family history of cancer was reported in 63% of cases, with the mother being the most frequently reported affected first-degree relative.

In terms of diagnostic pathways, 65% of patients were diagnosed following the onset of clinical symptoms, whereas only 35% were identified through routine screening or early detection programs.

With respect to disease stage at diagnosis, 22% of patients were diagnosed at Stage I, 34% at Stage II, and 44% at Stage III.

Table 1. Independent factors associated with advanced-stage breast cancer diagnosis (Stage III): results of multivariate logistic regression analysis.

Factor (Independent Variable)	**Reference Category**	**AOR**	**95% CIs**	***p*-value**
Nulliparity (no children)	Having children	6.42	2.87–14.36	<0.001
Symptom-based diagnosis	Screening-based diagnosis	4.91	2.24–10.77	<0.001
Physical inactivity	Regular physical activity	3.15	1.45–6.84	0.004
Positive family history of cancer	No family history	2.78	1.33–5.82	0.007
Vitamin D deficiency (<20 ng/mL)	Normal level (≥20 ng/mL)	2.21	1.12–4.35	0.022
Elevated fasting blood glucose (≥126 mg/dL)	Normal level (<100 mg/dL)	1.85	0.92–3.72	0.084

Table 1 presents the results of the multivariate logistic regression analysis identifying factors independently associated with advanced-stage breast cancer diagnosis (stage III). After adjustment for potential confounders, nulliparity emerged as the strongest predictor of late-stage diagnosis, with women who had no children being more than six times more likely to be diagnosed at stage III compared with those who had children (AOR = 6.42, 95% CI: 2.87–14.36; $p < 0.001$).

Mode of detection was also a significant determinant of disease stage. Women diagnosed following the onset of symptoms had nearly a fivefold higher likelihood of advanced-stage presentation compared with those diagnosed through routine screening (AOR = 4.91, 95% CI: 2.24–10.77; $p < 0.001$). In addition, physical inactivity was significantly associated with late-stage diagnosis, with

inactive women exhibiting more than three times higher odds of stage III disease compared with physically active women (AOR = 3.15, 95% CI: 1.45–6.84; $p = 0.004$). This finding is consistent with the results reported by Agodirin et al. (2021), who, in a systematic review and meta-analysis, identified similar determinants of delayed presentation and advanced-stage breast cancer diagnosis across African populations [15].

A positive family history of cancer was also independently associated with advanced-stage diagnosis (AOR = 2.78, 95% CI: 1.33–5.82; $p = 0.007$). Furthermore, women with vitamin D deficiency (<20 ng/mL) had more than twice the odds of being diagnosed at stage III compared with those with normal vitamin D levels (AOR = 2.21, 95% CI: 1.12–4.35; $p = 0.022$). This finding is consistent with the results reported by Shaukat et al. (2017), who demonstrated a significant association between vitamin D deficiency and breast cancer in a study conducted among Pakistani women [16].

Although elevated fasting blood glucose levels (≥126 mg/dL) were associated with an increased likelihood of advanced-stage diagnosis, this association did not reach statistical significance (AOR = 1.85, 95% CI: 0.92–3.72; $p = 0.084$). This finding aligns with the results of Jacobson et al. (2021), who reported hyperglycemia as an important metabolic risk factor for cancer development, although stage-specific associations were not consistently statistically significant, possibly reflecting differences in study design or limited statistical power [17].

At the initial step of the logistic regression model containing only the intercept, Age was not included in the equation; however, the Score test indicated that Age was significantly associated with the outcome. Specifically, the score statistic for Age was 5.455 with 1 degree of freedom, yielding a p-value of 0.020. This result suggests that Age is a statistically significant predictor of advanced-stage breast cancer diagnosis when considered individually, and therefore meets the criterion for inclusion in the multivariate logistic regression model.

Overall, these findings highlight the significant role of reproductive history, health-seeking behavior, lifestyle factors, familial risk, and selected metabolic markers in influencing the stage at which breast cancer is diagnosed in the Libyan population.

This study conducted a multivariate statistical analysis of the factors correlated with the stage of breast cancer diagnosis in a Libyan cohort. The high number of cases diagnosed in stage three (44%) shows that there is a systematic problem with late diagnosis, which is in line with reports from other developing nations [8]. The primary strength of this research resides in the application of a logistic regression model, which facilitated the isolation of the independent effects of diverse and overlapping risk factors.

The absence of children was the most significant independent predictor of advanced-stage diagnosis (AOR = 6.42). This phenomenon can be elucidated by the protective influence of pregnancy and lactation on breast tissues, which induce cellular modifications and culminate in cellular

differentiation, hence diminishing the likelihood of malignant transformation [9]. Additionally, women who have not given birth may have diminished engagement with the maternal and child health care system, hence decreasing the likelihood of inadvertent discovery.

The substantial link between symptom-based diagnosis and advanced stage (AOR = 4.91) shows that the main cause of the late diagnosis problem is the lack of structured screening programs. In Libya, diagnosis is mostly based on what the patient does once they show symptoms, which usually means the disease is already advanced [10]. This is different from countries that have national screening programs.

Our model also found that not getting enough exercise and not getting enough vitamin D are two risk variables that can be changed. Exercise can help you control your weight and hormone levels, and it can also reduce inflammation [11]. Vitamin D's involvement in regulating the cell cycle, differentiation, and decreasing tumor angiogenesis positions it as a possible biological element for prevention [12]. These findings create opportunities for non-clinical therapies.

Having a family history of cancer was a big deal since it showed that genetics and the environment were both involved. This shows how important it is to focus on women with a family history in intensive awareness and screening programs.

Acknowledgements

The authors would like to thank the administration and the Oncology Department of Al-Riaya Medical Clinic for facilitating data collection. The authors also express their sincere gratitude to all patients who participated in this study.

Conflict of Interest

The authors declare that there are no conflicts of interest associated with this research.

Data and Code Availability

The aggregated data and the statistical analysis syntax used in SPSS are available from the corresponding author upon reasonable request.

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

2. Bodalal, Z., Azzuz, R., & Bendardaf, R. (2019). Cancer in Libya: A systematic review. *Libyan Journal of Medicine, 14*(1), 1632671.

3. Yedjou, C. G., Sims, J. N., Miele, L., Noubissi, F., Lowe, L., Fonseca, D. D., ... & Tchounwou, P. B. (2019). Health and racial disparity in breast cancer. *Advances in Experimental Medicine and Biology, 1152*, 31–49.

4. Boder, J. M. E., Elmabrouk Abdalla, F. B., Elfageih, M. A., Abusaa, A., Buhmeida, A., & Collan, Y. (2011). Breast cancer patients in Libya: comparison with European and central African patients. *Oncology letters*, *2*(2), 323-330.

5. Naja, F., Nasreddine, L., Awada, S., El Sayed Ahmad, R., & Hwalla, N. (2019). Nutrition in the prevention of breast Cancer: a middle eastern perspective. *Frontiers in public health*, *7*, 316.

6. Azadnajafabad, S., Moghaddam, S. S., Mohammadi, E., Rezaei, N., Rashidi, M. M., Rezaei, N., ... & GBD 2019 NAME Breast Cancer Collaborators. (2023). Burden of breast cancer and attributable risk factors in the North Africa and Middle East region, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Frontiers in Oncology*, *13*, 1132816.

7. Lacey Jr, J. V., Devesa, S. S., & Brinton, L. A. (2002). Recent trends in breast cancer incidence and mortality. *Environmental and molecular mutagenesis*, *39*(2-3), 82-88.

8. Desai, K., Baralo, B., Kulkarni, A., Keshava, V. E., Iqbal, S., Ali, H., ... & Thirumaran, R. (2024). Cancer statistics: The United States vs. worldwide.

9. Coughlin, S. S. (2019). Epidemiology of breast cancer in women. *Advances in Experimental Medicine and Biology, 1152*, 9–29.

10. Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., ... & Zhu, H. P. (2017). Risk factors and preventions of breast cancer. *International Journal of Biological Sciences*, *13*(11), 1387–1397.

11. Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons.

12. El Mistiri, M., Verdecchia, A., Rashid, I., El Sahli, N., El Mangush, M., & Federico, M. (2007). Cancer incidence in eastern Libya: the first report from the Benghazi Cancer Registry, 2003. *International Journal of Cancer*, *120*(2), 392–397.

13. Alam, K., Maity, A., Sinha, S. K., Rizopoulos, D., & Sattar, A. (2021). Joint modeling of longitudinal continuous, longitudinal ordinal, and time-to-event outcomes. *Lifetime Data Analysis*, *27*(1).

14. Mertler, C. A., Vannatta, R. A., & LaVenia, K. N. (2021). *Advanced and multivariate statistical methods: Practical application and interpretation*. Routledge.

15. Agodirin, O., Aremu, I., Rahman, G., Olatoke, S., Olaogun, J., Akande, H., & Romanoff, A. (2021). Determinants of delayed presentation and advanced-stage diagnosis of breast cancer in Africa: a systematic review and meta-analysis. *Asian Pacific journal of cancer prevention: APJCP*, *22*(4), 1007.

16. Shaukat, N., Jaleel, F., Moosa, F. A., & Qureshi, N. A. (2017). Association between vitamin D deficiency and breast cancer. *Pakistan journal of medical sciences*, *33*(3), 645.

17. Jacobson, S., Dahlqvist, P., Johansson, M., Svensson, J., Billing, O., Sund, M., & Franklin, O. (2021). Hyperglycemia as a risk factor in pancreatic cancer: A nested case-control study using prediagnostic blood glucose levels. *Pancreatology*, *21*(6), 1112-1118.