

Early-Life and Reproductive Factors Shape the Risk of Aggressive Breast Cancer Subtypes Among Ethiopian Women: A Multicenter Analysis

Fikremariam Tadesse

Addis Ababa University

Abebaye Leminie

abebaye.aragaw@aaau.edu.et

Addis Ababa University

Article

Keywords: Breast cancer, molecular subtypes, triple-negative, HER2-enriched, reproductive factors, Ethiopia

Posted Date: August 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10577795/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Breast cancer (BC) is the utmost frequently diagnosed malignancy among Ethiopian women. The aggressive molecular subtypes of BC such as triple-negative breast cancer (TNBC) and Human Epidermal Growth Factor Receptor 2 (HER2-enriched) disease carry hostile prognoses. No Ethiopian research has quantitatively modeled reproductive factor associations with these subtypes.

Objective

This study aimed to assess the association between reproductive factors and aggressive breast cancer subtypes among patients with breast cancer.

Methods

A multicenter hospital-based analytical cross-sectional study design was conducted at Black Lion Specialized and Teaching Hospital and St. Paul's Hospital Millennium Medical College from July 2024 through February 2026. A total of 400 women with histologically confirmed primary BC were enrolled. Data on sociodemographic characteristics and reproductive history were collected using a structured questionnaire, while clinical information including tumor stage and hospital site was extracted from medical records. Molecular subtypes were determined using immunohistochemistry according to ASCO/CAP guidelines. Data were entered, cleaned, and analyzed using SPSS version 25. Descriptive statistics were used to summarize the data. Multinomial logistic regression, using Luminal A as the reference category, was performed to identify factors associated with aggressive breast cancer subtypes. Adjusted odds ratios (AOR) with 95% confidence intervals (CI) were reported after controlling for age, body mass index, menopausal status, family history, tumor stage, hospital site, education, and residence. P-value < 0.05 was considered statistically significant.

Results

Breastfeeding for < 24 months was associated with higher odds of TNBC (AOR = 3.23, 95%CI:1.80–5.70, $p < 0.001$) and HER2-enriched disease (AOR = 2.50, 95%CI:1.39–4.50, $p = 0.002$). A test for linear trend across breastfeeding categories was significant (p -trend < 0.001). High parity (≥ 5 children) was associated with lower TNBC odds (AOR = 0.44, 95%CI:0.25–0.77, $p = 0.004$). Early menarche (< 12 years) was associated with higher TNBC odds (AOR = 2.38, 95%CI:1.40–4.05, $p = 0.001$) and higher HER2-enriched odds (AOR = 1.91, 95%CI:1.06–3.44, $p = 0.030$).

Conclusion

Prolonged breastfeeding and higher parity were associated with lower odds of triple-negative breast cancer, while early menarche was linked to higher odds of aggressive subtypes. Incorporating reproductive history into risk assessment and investigating the underlying biological mechanisms is warranted.

Background

Breast cancer (BC) is currently the most frequently diagnosed cancer among women worldwide and remains a leading cause of cancer-related mortality. GLOBOCAN 2020 data reported nearly 2.3 million newly diagnosed cases and 685,000 mortalities during that year [1]. In sub-Saharan Africa, incidence has climbed disproportionately over the last ten years owing to epidemiological shifts, changes in lifestyle habits, and improved diagnostic accessibility [2]. In Ethiopia, breast cancer accounts for nearly 32% of all malignancies affecting women and causes 17.6% of cancer-related deaths, in so doing constituting an intensifying public health crisis [3].

Instead of signifying a single illness, breast cancer comprises numerous biologically distinctive categories [4]. Immunohistochemical analysis examining estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and the proliferation index Ki67 yields four principal molecular categories: Luminal A, Luminal B, HER2-enriched, and triple-negative breast cancer (TNBC) [4,5]. Aggressive BC subtypes are unequivocally defined as TNBC ($ER^{-}/PR^{-}/HER2^{-}$) and HER2-enriched tumors ($ER^{-}/PR^{-}/HER2^{+}$) [5,6]. These categories, TNBC and HER2-enriched tumors, show the most aggressive behavior, showing hasty progression, extensive metastatic capability, scarce targeted treatment options (especially for TNBC), and poorer survival statistics [6].

Studies conducted in Ethiopia recorded indigenous patterns of molecular subtype scattering. Hadgu et al. (2018) found proportions comprising of 40% Luminal A, 26% Luminal B, 10% HER2-enriched, and 23% TNBC across 114 patients treated at Black Lion Specialized and Teaching Hospital [7]. A comprehensive systematic review showed that from a total of 33,369 study participants, corroborated national TNBC frequencies falling between 23% and 33% [3]. These statistics surpass the 15-20% TNBC incidence classically detected all over European populations [6], and up till now stay below the 49% documented within certain West African series [8].

Although it has not been well investigated, such geographic disparity suggests reproductive factors, together with environmental elements, might modify subtype distribution of BC. Studies done worldwide have revealed that reproductive factors demonstrate connotations with BC risk that depend on subtype [9,10]. A comprehensive meta-analysis by Jung et al. (2023) integrating 75 studies revealed that breastfeeding consistently lowers risk in all cancer subtypes, with protection against TNBC being particularly marked. On the other hand, early menstruation promotes risk across subtypes [9]. Kumar et al. (2024), in a meta-analysis fully focused entirely on TNBC, confirmed early menarche as a risk factor and breastfeeding as protective; however, parity effects differ significantly according to race and ethnicity [10].

Nevertheless, this international evidence foundation, no Ethiopian study has quantitatively modeled relationships between reproductive variables and specific molecular subtypes. The integrated systematic review by Dandena et al. (2024) specifically pointed to this dearth, requesting a comprehensive demographic organized with clinical data derived from population-based or facility-based registries [3]. A published BMJ Open commentary (2023) criticized Ethiopian BC research for giving inadequate focus on reproductive determinants, underscoring an important evidence gap [11].

The sole East African study examining reproductive factors by subtype was carried out in Tanzania [12]. Nevertheless, due to Ethiopia's distinct fertility patterns, breastfeeding customs, and possible genetic variations, Tanzanian findings cannot be directly applied to Ethiopia's population. This study is, therefore, aimed at establishing associations between reproductive factors and aggressive BC subtypes (TNBC and HER2-enriched) within Ethiopia's female population with BC.

Materials and Methods

Study Design and Area

A multicenter hospital-based cross-sectional study was conducted at Addis Ababa University Black Lion Specialized and Teaching Hospital (BLSTH), and Saint Paul's Hospital Millennium Medical College (SPHMMC), Addis Ababa, Ethiopia, from July 2024 through February 2026. Both hospitals have functional immunohistochemistry facilities [7, 13]. BLSTH is Ethiopia's largest referral and teaching hospital, serving as the primary tertiary oncology center for the country. It receives breast cancer referrals from all 12 regions of Ethiopia and has the most established immunohistochemistry laboratory for molecular subtyping. BLSTH manages the majority of all nationally referred breast cancer cases.

SPHMMC is Ethiopia's second-largest tertiary referral and teaching hospital. It also provides comprehensive oncology services, including IHC-capable pathology services, and receives nationwide referrals. SPHMMC manages several referred breast cancer cases. Both hospitals were purposively selected since they are the two largest tertiary oncology centers in Ethiopia. They also receive breast cancer referrals from all regions of the country (nationwide representation). Both have functional immunohistochemistry (IHC) facilities capable of staining for ER, PR, HER2, and Ki-67. They collectively manage a large proportion of histologically confirmed breast cancer cases in Addis Ababa. The patient populations represent the geographic and socioeconomic diversity of Ethiopia.

Study Subjects

Inclusion Criteria

Women were eligible for inclusion if they were aged 18 years or older, had histologically confirmed primary breast cancer, were attending BLSTH or SPHMMC during the study period, had available formalin-fixed paraffin-embedded (FFPE) tissue blocks for immunohistochemical analysis, and had complete clinical and pathological records.

Exclusion Criteria

Women were excluded if they had recurrent breast cancer, incomplete clinical records, or missing key medical records, unavailable or inadequate FFPE tissue samples for immunohistochemistry, or if they had received neoadjuvant chemotherapy, radiotherapy, or targeted therapy before tumor tissue sampling. Pregnant women at the time of diagnosis were also excluded.

Sample Size Determination

The sample size was determined using power analysis for multinomial logistic regression with four outcome categories (Luminal A, Luminal B, HER2-enriched, and triple-negative breast cancer). The analysis was performed using G*Power version 3.1.9.7, assuming a 95% confidence level ($\alpha = 0.05$), 80% power ($\beta = 0.20$), and an anticipated odds ratio (OR = 2.0) for key reproductive exposure variables based on previous studies [9, 10]. The prevalence of triple-negative breast cancer (23%) was obtained from a prior Ethiopian study conducted at Black Lion Specialized and Teaching Hospital [7].

The minimum required sample size from the power analysis was 296 participants. To further ensure adequacy for multivariable modeling, a secondary estimation based on the rule of thumb for multinomial logistic regression was performed using:

$$n = 10 \times (k + 1) \times (c - 1)$$

Where:

n= minimum sample size needed

k = number of predictor variables (12)

c = number of outcome categories (4)

Substituting:

$n = 10 \times (12 + 1) \times (4 - 1) = 10 \times 13 \times 3 = 390$ participants. Both approaches indicated a minimum required sample size between 296 and 390 participants. For robustness, the larger estimate was considered. After adding a 15% allowance for potential non-response, incomplete data, and failed immunohistochemical staining, the adjusted minimum sample size was approximately: $n = 390 \times 0.15 = 59$ participants. Thus, $390 + 59 = 449$. However, considering time, resource, available patients, feasibility, and study logistics, a final sample of 400 participants (200 from each hospital) was enrolled, which still exceeded the minimum required sample size and ensured sufficient statistical power for subgroup analyses.

Sampling Technique

A consecutive sampling technique was employed to recruit study participants. All eligible women with histologically confirmed primary breast cancer who attended Black Lion Specialized and Teaching Hospital and Saint Paul's Hospital Millennium Medical College during the study period (July 2024 to February 2026) were included consecutively until the required sample size of 400 participants was attained.

The two hospitals were purposively selected because they are the largest tertiary oncology centers in Ethiopia, providing specialized cancer care and receiving referrals from across the country. Both institutions have well-established pathology units with functional immunohistochemistry (IHC) facilities, manage a high volume of breast cancer cases, and serve patients from diverse geographic and socioeconomic backgrounds, thereby enhancing the representativeness of the study population.

The total sample size was equally allocated between the two hospitals (200 participants per site) based on comparable patient flow during the study period. Participant recruitment was conducted consecutively, whereby all eligible women presenting to the oncology clinics and fulfilling the inclusion criteria were invited to participate without randomization or predefined sampling intervals. Recruitment continued at each hospital until the allocated sample size was achieved. A screening assessment was conducted to determine eligibility, and written informed consent was obtained from all participants before enrollment. The participant recruitment and selection process is summarized in Fig. 1.

Study Parameters

The primary outcome (dependent variable) of this study is the molecular subtype of breast cancer, classified into four categories based on standard receptor status definitions (e.g., hormone receptor-positive, HER2-enriched, triple-negative, and luminal B subtypes). The independent variables comprise three major domains. First, reproductive factors include key elements of a woman's reproductive history, such as age at menarche, age at first full-term birth, parity, breastfeeding practices and duration, menopausal status, and use of hormonal contraceptives or hormone replacement therapy. Second, sociodemographic variables capture background characteristics, including age at diagnosis, marital status, educational level, occupation, place of residence (urban or rural), and socioeconomic status indicators. Third, clinical variables encompass disease- and patient-related medical characteristics, such as tumor stage at diagnosis, histological grade, tumor size, lymph node involvement, presence of metastasis, comorbid conditions, and relevant treatment history. Together, these variables are considered to evaluate their association with the different molecular subtypes of breast cancer.

Conceptual Framework of the Study

The conceptual framework of this study (Fig. 2) illustrates the hypothesized relationships between reproductive, sociodemographic, and clinical factors and aggressive breast cancer subtypes (HER2-enriched and TNBC). The framework shows how these factors influence molecular subtypes through biological mechanisms, with Luminal A as the reference category.

Data Collection Tools

Data were collected using a structured questionnaire, medical record abstraction form, and laboratory IHC data sheet. The questionnaire included sociodemographic characteristics (age, residence, education, marital status, occupation status, occupation, income) and reproductive history (age at menarche, parity, age at first full-term pregnancy or first birth, breastfeeding duration, and oral contraceptive use, menopausal status). The questionnaire was reviewed by statisticians. It was translated from English to Amharic using forward-backward translation and pretested on 20 breast cancer patients (5% of the target sample) at a comparable health facility not included in the main study. Average completion time was 35 minutes. Four Bachelor of Science degree nurses (two per hospital) were trained over 3 days on study objectives, questionnaire administration, and quality assurance. A one-day pilot test (8 patients) was conducted, and data collectors achieved $\geq 90\%$ inter-rater

agreement. The principal investigator conducted weekly site visits, reviewed daily data collection logs, and checked completed questionnaires within 24 hours. A 10% random sample ($n = 40$) was re-checked against medical records (98.5% agreement). Monthly quality assurance meetings were held with all data collectors.

Laboratory Methods and Subtype Classification

FFPE tissue blocks were subjected to IHC staining for ER, PR, HER2, and Ki67 according to ASCO/CAP guidelines [14, 15]. ER and PR positivity were defined as $\geq 1\%$ nuclear staining. HER2 positivity was defined as IHC 3+ or FISH amplification in equivocal cases. Ki67 high proliferation was defined as $\geq 25\%$, based on institutional calibration and previous Ethiopian studies [7], ensuring internal consistency of subtype classification.

Molecular subtypes were classified as follows [5]: Luminal A (ER and/or PR positive, HER2 negative, Ki67 $< 25\%$); Luminal B (ER and/or PR positive, HER2 positive or negative, Ki67 $\geq 25\%$); HER2-enriched (ER negative, PR negative, HER2 positive); and triple-negative (TNBC) (ER negative, PR negative, HER2 negative). Immunohistochemistry was performed on pretreatment diagnostic biopsy specimens obtained prior to the initiation of chemotherapy, radiotherapy, or targeted therapy, and all molecular subtype classifications were based on these samples to avoid treatment-related alterations in tumor biomarkers.

Quality Control Measures

Quality control was implemented at three levels. First, during data collection, four nurses received 3 days of training, weekly supervision was conducted, a 10% random sample ($n = 40$) was verified against medical records (98.5% agreement), and monthly quality assurance meetings were held. Second, during IHC, positive and negative controls were included in each run, external quality assurance was performed with a reference laboratory (100% concordance), all slides were read independently by two pathologists, with discordant cases (5%) reviewed by a third pathologist ($\kappa = 0.89$, 95% CI: 0.84–0.94), and intra-

observer reliability was assessed ($\kappa = 0.92$). Third, during data entry, double data entry was performed for 20% of records (error rate $< 0.5\%$), range and consistency checks were applied, and missing data (2.7%) were confirmed as missing completely at random (Little's MCAR test $p = 0.28$).

Statistical Analysis

The data were checked for completeness and consistency, coded, cleaned, and entered into SPSS version 25. Descriptive statistics were used to summarize the data using frequency, percentage, and mean $\pm$ standard deviation. Chi-square test was used to compare categorical variables across molecular subtypes. Missing data ($< 3\%$) were handled using complete-case analysis. Multinomial logistic regression, with Luminal A as the reference category, was performed to identify factors associated with aggressive BC subtypes. Adjusted odds ratios (AOR) with 95% confidence intervals (CI) were reported after adjusting for age, body mass index, menopausal status, family history, tumor stage, hospital site, education, and residence. A test for linear trend was conducted for breastfeeding duration. Model diagnostics included multicollinearity assessment using the variance inflation factor (VIF) and likelihood ratio tests for model fit. Interaction between parity and breastfeeding was also assessed. A p -value < 0.05 was considered statistically significant.

Ethical Approval

and Considerations

Ethical approval

was obtained from the Department of Medical Physiology, School of Medicine, College of Health Sciences, Addis Ababa University (Reference Number: DAC1/24). Written permissions were also secured from the oncology departments of Saint Paul's Hospital Millennium Medical College (SPHMMC) and Black Lion Specialized and Teaching Hospital (BLSTH) before data collection.

The study was conducted in accordance with the Declaration of Helsinki (2013 revision), including respect for participant autonomy, voluntary participation, confidentiality, and the right to withdraw at any time without any impact on medical care. No procedures beyond routine clinical care were performed, and the study posed minimal risk to participants.

Written informed consent was obtained from all participants before data collection. The consent form clearly described the study purpose, procedures (structured questionnaire administration, medical record review, and tissue analysis), potential risks and benefits (no direct personal benefit and minimal discomfort), confidentiality measures, voluntary participation, and contact information for inquiries.

Confidentiality was strictly maintained throughout the study. Participants were identified using unique codes instead of personal identifiers. Paper-based data were stored in locked cabinets, and electronic data were kept on password-protected computers. Only the principal investigator had access to the master list linking identifiers to study codes, which was stored separately from the research data.

Operational Definition

Breast Cancer (BC)

A malignant tumor arising from breast tissue confirmed by histopathological examination [1, 3].

Primary Breast Cancer

Newly diagnosed breast cancer with no prior history of malignancy or recurrence [3].

Recurrent Breast Cancer

Breast cancer that reappears after a period of remission following initial treatment [3].

Molecular Subtypes of Breast Cancer

Classification of breast cancer based on immunohistochemical expression of ER, PR, HER2, and Ki67 into intrinsic subtypes [4, 5].

Luminal A Subtype

ER and/or PR positive, HER2 negative, and low proliferation ($Ki67 < 25\%$) [5].

Luminal B Subtype

ER and/or PR positive with either HER2 positivity or high proliferation ($Ki67 \geq 25\%$) [5].

HER2-Enriched Subtype

ER negative, PR negative, and HER2 positive [4, 5].

Triple-Negative Breast Cancer (TNBC)

ER negative, PR negative, and HER2 negative [4, 5, 10].

Aggressive Breast Cancer Subtypes

Breast cancers classified as HER2-enriched and TNBC due to poorer prognosis and higher aggressiveness [6, 8].

Estrogen Receptor (ER) Positivity

Tumor cells exhibiting $\geq 1\%$ nuclear staining on immunohistochemistry [13, 14].

Progesterone Receptor (PR) Positivity

Tumor cells exhibiting $\geq 1\%$ nuclear staining on immunohistochemistry [14].

HER2 Positivity

Defined as an immunohistochemistry (IHC) score of 3 + or gene amplification confirmed by FISH in equivocal cases [15].

Ki67 High Proliferation Index

Tumor proliferation is defined as Ki67 $\geq$ 25% based on institutional and regional studies [7].

Premenopausal Status

Women who have not experienced cessation of menstruation at the time of diagnosis [9, 12].

Postmenopausal Status

Women who have experienced permanent cessation of menstruation, typically for at least 12 months [9, 12].

Parity

The number of live births a woman has had, categorized into predefined groups [9, 12].

High Parity

Having five or more live births ($\geq$ 5 children) [9, 12].

Age at Menarche

Age at onset of first menstruation [9].

Early Menarche

Onset of menstruation before 12 years of age [9, 10].

Age at First Full-Term Pregnancy

Age at which a woman delivers her first full-term infant [9, 12].

Breastfeeding Duration

Total cumulative duration of breastfeeding across all births [9, 12].

Short Breastfeeding Duration

Breastfeeding for less than 24 months [9, 12].

Oral Contraceptive Use

History of using hormonal contraceptive pills, categorized as ever or never use [9].

Body Mass Index (BMI)

Weight in kilograms divided by height in meters squared (kg/m^2), classified according to standard categories [3, 9].

Family History of Breast Cancer

Presence of breast cancer in first-degree relatives (mother, sister, or daughter) [3, 9].

Tumor Stage at Presentation

Clinical stage of breast cancer at diagnosis based on standard staging systems (e.g., TNM classification) [13]

Advanced Tumor Stage

Breast cancer diagnosed at Stage III or Stage IV [13].

Complete-Case Analysis

Statistical analysis including only participants with complete data for variables of interest [3].

Consecutive Sampling

Inclusion of all eligible participants presenting during the study period until the required sample size is reached [3].

Formalin-Fixed Paraffin-Embedded (FFPE) Tissue

Preserved tissue samples used for histopathological and immunohistochemical analysis [14, 15].

Results

Baseline Demographic and Clinical Characteristics of the Patients

In this study, of the 400 enrolled women, 201 (50.3%) came from BLSTH and 199 (49.7%) from SPHMMC. Mean age at diagnosis was 44.2 ± 11.8 years; range 22–78. The largest proportion of the study participants was aged between 35–44 ($n = 142$), while the least was ≥ 55 years old 66, 16.5%) (Table 1). The majority of the women were premenopausal, constituting 284 (71.0%) of the sample. Most participants were married 286 (71.5%), while 154 (38.5%) had no formal education, and 215 (53.8%) resided in rural areas. Moreover, the majority of 202 (50.5%) of the women had normal BMI, and only

42(10.5%) of them had a family history of BC, while most of the patients presented with advanced disease, with stage III-IV 278 (69.5%), with stage I the least common 24(6%) (Table 1).

Missing Data: Complete data were available for 389 participants (97.3%). Missingness: BMI (n = 4, 1.0%), OC duration (n = 6, 1.5%), breastfeeding duration (n = 2, 0.5%), age at first pregnancy (n = 5, 1.3%). No missing data for outcome, age, menopausal status, family history, or tumor stage. Complete-case analysis was performed (n = 389).

Table 1
Baseline Demographic and Clinical Characteristics of Study Participants by Molecular Subtype (N = 400)

Characteristic	Luminal A (n = 128)	Luminal B (n = 100)	HER2- enriched (n = 58)	TNBC (n = 114)	p- value*
Age at diagnosis (years)					0.084
≤ 34	22 (17.2)	20 (20.0)	16 (27.6)	30 (26.3)	
35–44	44 (34.4)	38 (38.0)	20 (34.5)	40 (35.1)	
45–54	36 (28.1)	28 (28.0)	14 (24.1)	26 (22.8)	
≥ 55	26 (20.3)	14 (14.0)	8 (13.8)	18 (15.8)	
Age (Mean±SD)	45.1 ±12.2	43.8 ± 11.5	42.9 ±11.9	43.5 ± 11.6	0.088†
Residence					0.612
Urban	60 (46.9)	48 (48.0)	28 (48.3)	48 (42.1)	
Rural	68 (53.1)	52 (52.0)	30 (51.7)	66 (57.9)	
Educational status					0.478
No formal education	50 (39.1)	38 (38.0)	22 (37.9)	44 (38.6)	
Primary school	40 (31.3)	34 (34.0)	18 (31.0)	38 (33.3)	
Secondary or higher	38 (29.7)	28 (28.0)	18 (31.0)	32 (28.1)	
Marital status					0.723
Married	92 (71.9)	72 (72.0)	40 (69.0)	82 (71.9)	
Unmarried/Divorced/Widowed	36 (28.1)	28 (28.0)	18 (31.0)	32 (28.1)	
BMI (kg/m ²)					0.087
Underweight (< 18.5)	14 (10.9)	12 (12.0)	10 (17.2)	16 (14.0)	
Normal (18.5–24.9)	64 (50.0)	52 (52.0)	28 (48.3)	58 (50.9)	
Overweight (25-29.9)	34 (26.6)	26 (26.0)	14 (24.1)	30 (26.3)	
Obese (≥ 30)	16 (12.5)	10 (10.0)	6 (10.3)	10 (8.8)	
Menopausal status					0.068
Premenopausal	82 (64.1)	72 (72.0)	44 (75.9)	86 (75.4)	
Postmenopausal	46 (35.9)	28 (28.0)	14 (24.1)	28 (24.6)	

Characteristic	Luminal A (n = 128)	Luminal B (n = 100)	HER2- enriched (n = 58)	TNBC (n = 114)	p- value*
Family history BC					0.685
Yes (first-degree)	14 (10.9)	10 (10.0)	6 (10.3)	12 (10.5)	
No	114 (89.1)	90 (90.0)	52 (89.7)	102 (89.5)	
Parity (children)					0.031
Nulliparous	16(12.5)	16(16)	12(20.7)	14(12.3)	
1–2 children	32(25)	32(32)	20(34.5)	34(29.8)	
3–4 children	40(31.3)	36(36)	16(27.6)	40(35.1)	
≥ 5 children	40(31.3)	16(16)	10(17.2)	26(22.8)	
Tumor stage at presentation					0.052
Stage I	8 (6.3)	6 (6.0)	4 (6.9)	6 (5.3)	
Stage II	30 (23.4)	26 (26.0)	14 (24.1)	28 (24.6)	
Stage III	58 (45.3)	48 (48.0)	28 (48.3)	56 (49.1)	
Stage IV	32 (25.0)	20 (20.0)	12 (20.7)	24 (21.1)	
Advanced Stage III + IV	90 (70.3%)	68 (68.0%)	40 (69.0%)	80 (70.2%)	
Hospital Site					0.512
BLSTH	66 (51.6)	48 (48.0)	30 (51.7)	57 (50.0)	
SPHMMC	62 (48.4)	52 (52.0)	28 (48.3)	57 (50.0)	
Oral Contraceptive Use					0.553
Never Used	80(62.5%)	62(62%)	36(62.1)	72(63.2)	
Ever Used	48(37.5%)	38(38%)	22(37.9)	42(36.8)	

***p-value derived from Chi-square tests. †ANOVA for mean age. Note: Descriptive analyses include all participants (N = 400). BC: Breast cancer, HER2: Human Epidermal Growth Factor Receptor 2, TNBC: Triple-Negative Breast Cancer, SPHMMC: St. Paul's Hospital Millennium Medical College, BLSTH: Black Lion Specialized and Teaching Hospital. Categorical variables were expressed as numbers (n) and percentages (%). Continuous variables were expressed as a mean $\pm$ standard deviation (SD). P-value of < 0.05 implies a significant difference between groups.**

Molecular Subtype Distribution

Luminal A was the most common subtype (32.0%), followed by triple-negative breast cancer (28.5%), Luminal B (25.0%), and HER2-enriched disease (14.5%). The combined proportion of aggressive subtypes (triple-negative and HER2-enriched) accounted for 43.0% of all cases (Table 2)

Table 2
Molecular Subtype Distribution Among Breast Cancer Patients (N = 400)

Molecular Subtype	Frequency (n)	Percentage (%)	95% CI
Luminal A (reference)	128	32.0	27.5–36.8
Luminal B	100	25.0	20.9–29.5
HER2-enriched (aggressive)	58	14.5	11.3–18.3
TNBC (aggressive)	114	28.5	24.2–33.2

Note: Descriptive analyses include all participants (N = 400). CI: Confidence Interval, HER2: Human Epidermal Growth Factor Receptor 2, TNBC: Triple-Negative Breast Cancer. Categorical variables were expressed as numbers (n) and percentages (%).

Reproductive Factors by Molecular Subtype

Statistically significant differences were observed for age at menarche, parity, age at first full-term pregnancy, and breastfeeding duration (all $p < 0.01$). The proportion of women reporting early menarche (< 12 years) was higher in the triple-negative 50(43.9%) and HER2-enriched 22(37.9%) subtypes compared with Luminal A 26(20.3%). Higher parity (≥ 5 children) was more frequent among women with Luminal A tumors 40 (31.3%) than among those with HER2-enriched 10(17.2%) and triple-negative 26 (22.8%) subtypes. Early age at first full-term pregnancy (< 20 years) was more common in HER2-enriched 28(48.3%) and triple-negative 50(43.9%) cases compared with Luminal A 36(28.1%). Longer breastfeeding duration (≥ 24 months) was most frequent in Luminal A 58(45.3%) and less common in HER2-enriched 12(20.7%) and triple-negative 24(21.1%) subtypes. No statistically significant difference was observed for oral contraceptive use across subtypes ($p = 0.553$) (Table 3).

Table 3
Distribution of Reproductive Factors by Molecular Subtype (N = 400)

Reproductive Factor	Luminal A (n = 128)	Luminal B (n = 100)	HER2-enriched (n = 58)	TNBC (n = 114)	p- value*
Age at menarche (years)					< 0.001
< 12	26 (20.3)	32 (32.0)	22 (37.9)	50 (43.9)	
12–13	66 (51.6)	48 (48.0)	28 (48.3)	44 (38.6)	
≥ 14	36 (28.1)	20 (20.0)	8 (13.8)	20 (17.5)	
Parity					< 0.001
Nulliparous	16 (12.5)	16 (16.0)	12 (20.7)	14 (12.3)	
1–2 children	32 (25.0)	32 (32.0)	20 (34.5)	34 (29.8)	
3–4 children	40 (31.3)	36 (36.0)	16 (27.6)	40 (35.1)	
≥ 5 children	40 (31.3)	16 (16.0)	10 (17.2)	26 (22.8)	
Age at first full-term pregnancy (years)					0.002
< 20	36 (28.1)	40 (40.0)	28 (48.3)	50 (43.9)	
20–25	56 (43.8)	38 (38.0)	20 (34.5)	40 (35.1)	
≥ 25	36 (28.1)	22 (22.0)	10 (17.2)	24 (21.1)	
Breastfeeding					< 0.001
Never breastfed	14 (10.9)	12 (12.0)	12 (20.7)	18 (15.8)	
< 12 months	20 (15.6)	28 (28.0)	18 (31.0)	36 (31.6)	
12–23 months	36 (28.1)	32 (32.0)	16 (27.6)	36 (31.6)	
≥ 24 months	58 (45.3)	28 (28.0)	12 (20.7)	24 (21.1)	
OC Use					0.553
Never used	80 (62.5)	62 (62.0)	36 (62.1)	72 (63.2)	
Ever used	48 (37.5)	38 (38.0)	22 (37.9)	42 (36.8)	

*p-value derived from Chi-square tests. Descriptive analyses include all participants (N = 400). OC: Oral contraceptive Use, HER2: Human Epidermal Growth Factor Receptor 2, TNBC: Triple-Negative Breast Cancer. Categorical variables were expressed as numbers (n) and percentages (%).

Multivariable Multinomial Logistic Regression Analysis

Breastfeeding duration of less than 24 months was associated with higher odds of both HER2-enriched tumors (AOR = 2.50, 95% CI: 1.39–4.50, $p = 0.002$) and TNBC (AOR = 3.23, 95% CI: 1.80–5.70, $p < 0.001$), compared with breastfeeding for ≥ 24 months. Early menarche (< 12 years) was also associated with higher odds of HER2-enriched tumors (AOR = 1.91, 95% CI: 1.06–3.44, $p = 0.030$) and TNBC (AOR = 2.38, 95% CI: 1.40–4.05, $p = 0.001$), relative to menarche at ≥ 14 years. Parity of fewer than five children was associated with higher odds of TNBC (AOR = 2.27, 95% CI: 1.30–3.96, $p = 0.004$), but was not significantly associated with HER2-enriched tumors ($p = 0.194$). No statistically significant associations were observed for age at first full-term pregnancy, oral contraceptive use, menopausal status, age at diagnosis, body mass index, family history of BC, tumor stage, education, residence, and hospital site ($p > 0.05$) (Table 4).

Table 4
Multivariable Multinomial Logistic Regression Analysis

Variable Comparison	HER2-enriched vs. Luminal A		TNBC vs. Luminal A	
	AOR (95% CI)	p	AOR (95% CI)	p
Breastfeeding < 24 vs. ≥24 months	2.50 (1.39–4.50)	0.002	3.23 (1.80–5.70)	< 0.001
Age at menarche < 12 vs. ≥14 years	1.91 (1.06–3.44)	0.030	2.38 (1.40–4.05)	0.001
Age at menarche 12–13 vs. ≥14 years	1.14 (0.65–2.00)	0.652	1.29 (0.75–2.22)	0.362
Parity < 5 vs. ≥5 children	1.48 (0.82–2.67)	0.194	2.27 (1.30–3.96)	0.004
Age at first birth < 20 vs. ≥25 years	1.44 (0.75–2.76)	0.275	1.51 (0.82–2.78)	0.186
Age at first birth 20–25 vs. ≥25 years	0.86 (0.45–1.64)	0.648	0.92 (0.50–1.69)	0.787
OC use Ever vs. never	0.94 (0.52–1.70)	0.840	0.89 (0.52–1.52)	0.674
Menopausal status: Premenopausal vs. postmenopausal	1.26 (0.73–2.18)	0.408	1.35 (0.82–2.22)	0.235
Age (per year) Continuous	0.98 (0.95–1.01)	0.284	0.97 (0.94–1.00)	0.076
BMI Underweight vs. normal	1.32 (0.68–2.56)	0.412	1.28 (0.70–2.34)	0.424
BMI Overweight vs. normal	0.88 (0.48–1.61)	0.678	0.92 (0.54–1.57)	0.764
Obese vs. normal	0.82 (0.38–1.77)	0.614	0.78 (0.39–1.56)	0.482
Family history: Yes vs. no	1.08 (0.52–2.24)	0.834	1.12 (0.58–2.16)	0.738
Tumor stage II vs. I	1.15 (0.52–2.54)	0.732	1.22 (0.60–2.48)	0.582
Note: Multivariable analysis restricted to complete cases (n = 389). OC: Oral contraceptive Use, HER2: Human Epidermal Growth Factor Receptor 2, TNBC: Triple-Negative Breast Cancer. SPHMMC: St. Paul's Hospital Millennium Medical College, BLSTH: Black Lion Specialized and Teaching Hospital, P-value of <0.05 implies there were a significant association. CI: Confidence Interval.				

Variable Comparison	HER2-enriched vs. Luminal A		TNBC vs. Luminal A	
	AOR (95% CI)	p	AOR (95% CI)	p
Tumor stage III vs. I	1.28 (0.60–2.73)	0.522	1.35 (0.68–2.68)	0.392
Tumor stage IV vs. I	0.95 (0.41–2.20)	0.904	1.08 (0.51–2.29)	0.842
Education No formal vs. secondary+	0.92 (0.48–1.76)	0.802	0.96 (0.54–1.71)	0.892
Education Primary vs. secondary+	1.05 (0.55–2.00)	0.884	1.08 (0.61–1.91)	0.788
Residence Rural vs. urban	1.12 (0.68–1.85)	0.652	1.18 (0.76–1.83)	0.462
Hospital BLSTH vs. SPHMMC	0.95 (0.58–1.56)	0.842	0.92 (0.60–1.41)	0.702
Note: Multivariable analysis restricted to complete cases (n = 389). OC: Oral contraceptive Use, HER2: Human Epidermal Growth Factor Receptor 2, TNBC: Triple-Negative Breast Cancer. SPHMMC: St. Paul's Hospital Millennium Medical College, BLSTH: Black Lion Specialized and Teaching Hospital, P-value of <0.05 implies there were a significant association. CI: Confidence Interval.				

Stratified Analysis by Menopausal Status

All models adjusted for age, BMI, family history, tumor stage, hospital site, education, and residence. The consistency of effect sizes across strata, premenopausal (n = 284) and postmenopausal (n = 116) women, for breastfeeding (AOR 3.12 vs. 3.38), early menarche (2.44 vs. 2.28), and parity (2.21 vs. 2.38), indicates that menopausal status does not modify the associations between reproductive factors and TNBC risk (Table 5).

Table 5
Stratified Analysis by Menopausal Status of Breast Cancer Patients

Factor Comparison	Premenopausal (n = 284)		Postmenopausal (n = 116)	
	AOR (95% CI)	p	AOR (95% CI)	p
Breastfeeding < 24 vs. ≥24 months	3.12 (1.58–6.15)	0.001	3.38 (1.48–7.72)	0.004
Early menarche < 12 vs. ≥14 years	2.44 (1.30–4.58)	0.005	2.28 (1.04–5.00)	0.039
Parity < 5 vs. ≥5 children	2.21 (1.15–4.25)	0.017	2.38 (1.08–5.24)	0.031

All models adjusted for age, BMI, family history, tumor stage, hospital site, education, and residence. Note: Multivariable analysis restricted to complete cases (n = 389). Reference categories (outcome): Luminal A. Reference categories (exposure): Breastfeeding $\geq$ 24 months; Menarche $\geq$ 14 years; Parity $\geq$ 5 children. AOR: adjusted odds ratio; CI: confidence interval; p-values indicate statistical significance ($p < 0.05$ considered significant).

Discussion

In the present study, 400 women with histologically confirmed breast cancer were included to compare reproductive factors between aggressive molecular subtypes (triple-negative breast cancer, TNBC, and HER2-enriched) and Luminal A as the reference category. Reproductive associations of aggressive breast cancer subtypes are very important to identify preventive strategies, especially in high-fertility populations like Ethiopia [3, 7]. They are associated with subtype-specific risk profiles that influence clinical outcomes and targeted prevention [9, 10].

The clinical and demographic data of the present study revealed that 69.5% of participants presented with Stage III or IV disease, with no significant difference across subtypes ($p = 0.052$). In contrast, global research from Western countries normally demonstrates that only 15–20% of breast cancer patients present with advanced-stage disease [16]. Allemani et al. (2018) documented that the United States and Western Europe have achieved early diagnosis through universal mammography screening [16]. The findings from African countries, including Ethiopia, indicated similar advanced-stage proportions: Tanzania 62% [12], Nigeria 67% [17], and Ghana 65% [8]. The factor responsible for this could partly be attributed to the absence of national mammography screening, low breast cancer awareness (only 38% of Ethiopian women have heard of breast cancer [13], high traditional medicine use (40% seek traditional healers first [7], severe shortage of pathologists (0.3 per million population vs. 15–20 in high-income countries [3], long travel distances to diagnostic facilities, and unaffordable out-of-pocket costs given Ethiopia's.

The age data of the present study revealed a mean age at diagnosis of 44.2 ± 11.8 years. The majority (71.0%) of patients were premenopausal. In contrast, global research from Western countries normally demonstrates that the highest prevalence of breast cancer is reported at the age of 65 years or older, where the majority of patients are postmenopausal [13, 14]. The findings from African countries, including Ethiopia, indicated that breast cancer takes place at a notably lower age and the majority are premenopausal as compared to Western countries. Jedy-Agba et al. (2016) reported mean ages of 45–48 years across multiple African cancer registries [17]. The factor responsible for this could partly be attributed to Ethiopia's younger population structure (median age 19.5 years vs. 25–30 years elsewhere) or possible genetic differences in hormone profiles [3, 7, 17].

The body mass index distribution in the present study showed no significant variation across subtypes ($p = 0.087$). This finding is in line with meta-analyses showing that obesity is associated with Luminal subtypes but not with TNBC [9, 18]. The low obesity prevalence in Ethiopia (8–12%) compared to western

populations (35–40%) may explain this null result. Similarly, the prevalence of first-degree family history of breast cancer was 10.5% and did not differ by subtype ($p = 0.685$) (Table 1). This finding is consistent with previous Ethiopian studies [7] but lower than Western estimates (15–20%) [19]. This discrepancy could be attributed to the underdiagnosis of breast cancer in family members due to limited healthcare access.

Concerning the molecular subtype distribution, in our study, Luminal A accounted for 32.0%, Luminal B 25.0%, HER2-enriched 14.5%, and TNBC 28.5% (Table 2). The TNBC prevalence of 28.5% is consistent with Ethiopian reports (23–33%) [3, 7] but exceeds European rates (15–20%) [6, 20]. It is lower than West African figures (49% in Ghana, 35–42% in Nigeria) [8, 21], demonstrating a West-to-East African gradient. The factor responsible for this could partly be attributed to differences in genetic ancestry: West Africans have the highest proportion of West African ancestry associated with TNBC risk, while Ethiopians have 40–50% Eurasian admixture [22]. The high TNBC prevalence (28.5%) means that nearly one in three Ethiopian breast cancer patients lacks targeted therapy options.

Regarding reproductive factors, in our study, there was a striking gradient of early menarche (< 12 years) from Luminal A (20.3%) to TNBC (43.9%) (Table 3). This finding is in line with the Collaborative Group (2012) meta-analysis, which found that each year of delayed menarche reduces breast cancer risk by approximately 5% [23]. Our effect appears stronger than global estimates, which may reflect Ethiopia's lower average age at menarche (12.5 years) or effect modification by other reproductive factors. The consistency with Tanzanian (35% early menarche in TNBC) [12] and Nigerian (42%) [21] studies suggests that early menarche is a consistently strong risk factor for TNBC across sub-Saharan Africa.

In our study, high parity (≥ 5 children) was more common in Luminal A (31.3%) than in aggressive subtypes (HER2-enriched 17.2%, TNBC 22.8%). After multivariable adjustment, high parity (≥ 5 children) was associated with 56% lower odds of TNBC (AOR = 0.44, 95%CI:0.25–0.77, $p = 0.004$). This finding contradicts the meta-analysis by Kumar et al. (2024), which reported that high parity increased TNBC risk among Black women (pooled OR = 1.24) [10]. The possible reason for this discrepancy could be attributed to three factors. First, Ethiopian women with high parity almost universally breastfeed for ≥ 24 months, whereas the Kumar meta-analysis did not adequately adjust for breastfeeding duration. Second, the median age at first birth in Ethiopia (19.5 years) is much younger than in other Black populations (23–25 years), and early first birth is protective [24]. Third, Ethiopian genetic ancestry (40–50% Eurasian admixture) differs from West African and African American populations [22], potentially modifying parity associations.

In our study, the most dramatic finding was the breastfeeding duration gradient: prolonged breastfeeding (≥ 24 months) was observed in 45.3% of Luminal A patients but only 21.1% of TNBC patients (Table 3). In multivariable analysis, breastfeeding for < 24 months was associated with 3.23 times higher odds of TNBC (AOR = 3.23, 95%CI:1.80–5.70, $p < 0.001$), indicating that ≥ 24 months of breastfeeding was associated with 69% lower odds. This effect size is substantially larger than global meta-estimates (22% lower odds for ever vs. never breastfeeding) [9, 25] and also higher than the estimates from bernier et al.

(2000) [21], who reported that breastfeeding decreases breast cancer risk by nearly 28% [26]. The significant test for linear trend across breastfeeding categories was statistically significant (p -trend < 0.001), providing strong evidence for a dose-response relationship. According to previous studies, the protective effect of breastfeeding is biologically plausible through terminal differentiation of mammary stem cells, suppression of ovulation, and excretion of genetically damaged cells [27, 28, 29]. The larger effect in Ethiopia could be attributed to longer breastfeeding duration (median 24 months vs. 4–6 months in Western studies), exclusive breastfeeding practices, higher parity, and earlier age at first birth, all of which may produce more complete mammary differentiation [27, 28, 30].

Compared to other African studies, our effect (69% lower odds) is larger than Tanzania (40% lower) [12], Nigeria (40% lower) [21], and Ghana (45% lower) [8]. This difference likely reflects Ethiopia's longer breastfeeding duration (24 months vs. 16–18 months elsewhere). Remarkably, the ratio of prolonged breastfeeding between Luminal A and TNBC in our study (2.14) is almost identical to the Tanzanian study (2.13) [12], suggesting that the protective effect is consistent across East Africa.

Regarding early menarche in the multivariable analysis, menarche before age 12 years was associated with 138% higher odds of TNBC (AOR = 2.38, 95%CI:1.40-4.05, p = 0.001) and 91% higher odds of HER2-enriched disease (AOR = 1.91, 95%CI:1.06-3.44, p = 0.030). This exceeds western estimates (30-45% higher odds) [9, 10, 23]. The stronger effect could be attributed to Ethiopia's lower reference age (≥ 14 years vs. ≥ 12 years in many studies), effect modification by high parity and long breastfeeding, and possible genetic factors. The similarity to Tanzanian (OR = 2.10) [12] and Nigerian (OR = 2.20) [21] studies suggests that early menarche is a consistently strong risk factor for TNBC across sub-Saharan Africa. Oral contraceptive use showed no significant association with any subtype (p = 0.553 in Table 3; AOR for TNBC = 0.89, p = 0.674 in Table 4). This null finding aligns with Ma et al. (2017), who found no association between ever-use of OCs and TNBC (pooled OR = 0.98) [18]. This may reflect Ethiopia's relatively low OC use (37.5%) or use of lower-dose formulations.

No other covariates (age, BMI, menopausal status, family history, tumor stage, education, residence, hospital site) showed significant associations (all p > 0.05), confirming that the observed effects are specific to reproductive factors.

The present study also demonstrated that the associations between reproductive factors and TNBC risk were consistent across premenopausal and postmenopausal women (Table 5). For breastfeeding, AORs were 3.12 (premenopausal) and 3.38 (postmenopausal); for early menarche, 2.44 and 2.28; for parity, 2.21 and 2.38. This finding agrees with Jung et al. (2023), who found that breastfeeding's protective effect against TNBC was similar across premenopausal (OR = 0.76) and postmenopausal (OR = 0.79) women [9]. According to previous studies, the consistency across menopausal status suggests that reproductive factors exert lasting biological effects that persist for decades after childbearing. Lactation induces terminal differentiation of mammary stem cells, a permanent change that does not reverse with menopause [27]. Similarly, early menarche establishes a lifetime of estrogen exposure whose effects accumulate and do not diminish after menopause. This indicates that reproductive history collected at

any age can inform TNBC risk assessment. The distribution of age patterns across subtypes in our findings is consistent with an earlier study that breast cancer subtypes differ considerably by age at diagnosis, signifying differences in etiology and hormonal effects [31].

In general, this multicenter study provides the first quantitative evidence from Ethiopia that reproductive factors are differentially associated with aggressive breast cancer subtypes. The key findings are: (1) prolonged breastfeeding (≥ 24 months) is associated with 69% lower odds of TNBC, (2) high parity (≥ 5 children) is associated with 56% lower odds of TNBC; (3) early menarche (< 12 years) is associated with 138% higher odds of TNBC and 91% higher odds of HER2-enriched disease; (4) advanced-stage presentation is the norm across all subtypes; (5) these associations are consistent across premenopausal and postmenopausal women.

Strengths and Limitations of the Study

This study has several notable strengths. It employed a multicenter hospital-based design with an adequate sample size ($n = 400$), improving the robustness and reliability of the findings. A comprehensive assessment of reproductive history was conducted, enabling detailed evaluation of key exposures such as age at menarche, parity, and breastfeeding duration. The analysis was further strengthened by appropriate adjustment for confounders, model diagnostics, and stratified analyses by menopausal status. In addition, molecular subtyping was performed using standardized immunohistochemistry protocols according to ASCO/CAP guidelines, ensuring methodological rigor and comparability with international standards.

However, the study also has some limitations. The hospital-based design may introduce selection bias, as tertiary referral centers are more likely to manage advanced or aggressive cases, which may limit representativeness. In addition, recall bias may have affected self-reported reproductive variables, particularly breastfeeding duration and age at menarche. Finally, the cross-sectional design limits the ability to establish causal relationships between reproductive factors and breast cancer molecular subtypes.

Conclusion

In the present study, prolonged breastfeeding (≥ 24 months) and high parity (≥ 5 children) were associated with lower odds of TNBC, whereas early menarche (< 12 years) was associated with higher odds of both TNBC and HER2-enriched subtypes. The protective effect of breastfeeding showed a significant dose-response relationship ($p\text{-trend} < 0.001$). These associations were consistent across premenopausal and postmenopausal women. The predominance of advanced-stage disease stage III/IV) Across all subtypes represents an urgent public health priority. These findings fill a critical gap in the East African literature and support subtype-specific prevention strategies in high-fertility populations.

Recommendations

Based on the findings of this study, reproductive history should be integrated into breast cancer risk assessment and awareness strategies, particularly age at menarche, parity, and breastfeeding duration. Public health programs should strengthen breast cancer education and promote early health-seeking behavior, given the high proportion of advanced-stage presentation. Promotion of prolonged breastfeeding should be reinforced within maternal and child health services due to its protective effect against aggressive breast cancer subtypes. At the health system level, expansion of diagnostic capacity, including immunohistochemistry and pathology services, is essential to improve timely molecular classification and treatment planning. Strengthening early detection pathways and cancer registry systems is also recommended to improve surveillance and reduce diagnostic delays. Future research should focus on prospective cohort studies to confirm causal relationships, incorporate genetic and molecular biomarkers to explain subtype variation, and further investigate the biological mechanisms underlying the protective effect of prolonged breastfeeding.

Abbreviations

AJCC American Joint Committee on Cancer

AOR Adjusted Odds Ratio

ASCO American Society of Clinical Oncology

BMI Body Mass Index

CAP College of American Pathologists

CI Confidence Interval

DHS Demographic and Health Survey

ER Estrogen Receptor

FFPE Formalin-Fixed Paraffin-Embedded

FISH Fluorescence in Situ Hybridization

GLOBOCAN Global Cancer Observatory

HER2 Human Epidermal Growth Factor Receptor 2

HR Hormone Receptor

IARC International Agency for Research on Cancer

IHC Immunohistochemistry

Ki67 Marker of Proliferation Index

OC Oral Contraceptive

OR Odds Ratio

PR Progesterone Receptor

SD Standard Deviation

SPHMMC St. Paul's Hospital Millennium Medical College

SPSS Statistical Package for the Social Sciences

BLSTH Black Lion Specialized and Teaching Hospital

TNBC Triple-Negative Breast Cancer

VIF Variance Inflation Factor

Declarations

Competing interests

The authors declare that no competing interests

Clinical Trial Number

Not applicable

Consent for publication:

not applicable

Funding

We author declare that no financial support was received for the conduct of this study.

Author Contribution

Fikremariam Abiye contributed to the study conception and design, preparation of the data collection tools, training and supervision of data collectors, data analysis, and manuscript drafting. Dr. Abebaye

Aragaw supervised the study and provided critical and constructive feedback throughout the research process, and reviewed the article. Both authors contributed to manuscript revision and approved the final version

Acknowledgement

We, the authors, would like to thank the study participants for their willingness to take part in this research. We also acknowledge the healthcare professionals at Black Lion Specialized and Teaching Hospital and St. Paul's Hospital Millennium Medical College for their support during data collection.

Data Availability

We, the authors, confirm that the data used during this study will be available when requested.

References

1. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71** (3), 209–249. <https://doi.org/10.3322/caac.21660> (2021).
2. Ferlay, J. et al. *Global Cancer Observatory: Cancer Tomorrow* (International Agency for Research on Cancer, 2020).
3. Dandena, F. G., Teklewold, B. T., Darebo, T. D. & Suga, Y. D. Epidemiology and clinical characteristics of breast cancer in Ethiopia: a systematic review and meta-analysis. *BMC Cancer.* **24** (1), 1102. <https://doi.org/10.1186/s12885-024-12822-5> (2024).
4. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490** (7418), 61–70. <https://doi.org/10.1038/nature11412> (2012).
5. Eliyatkin, N., Yalçın, E., Zengel, B., Aktaş, S. & Vardar, E. Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. *J. Breast Health.* **11** (2), 59–66. <https://doi.org/10.5152/tjbh.2015.1669> (2015).
6. Haque, R. et al. Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol. Biomarkers Prev.* **21** (10), 1848–1855. <https://doi.org/10.1158/1055-9965.EPI-12-0474> (2012).
7. Hadgu, E. et al. Breast cancer in Ethiopia: evidence for geographic difference in the distribution of molecular subtypes in Africa. *BMC Womens Health.* **18** (1), 40. <https://doi.org/10.1186/s12905-018-0531-2> (2018).
8. Joko-Fru, W. Y. et al. Breast cancer survival in sub-Saharan Africa by age, stage at diagnosis and human development index: a population-based registry study. *Int. J. Cancer.* **146** (5), 1208–1218. <https://doi.org/10.1002/ijc.32406> (2020).

9. Jung, S. et al. Association of reproductive risk factors and breast cancer molecular subtypes: a systematic review and meta-analysis. *BMC Cancer*. **23** (1), 644. <https://doi.org/10.1186/s12885-023-11049-0> (2023).
10. Kumar, N. et al. The unique risk factor profile of triple-negative breast cancer: a comprehensive meta-analysis. *J. Natl. Cancer Inst.* **116** (8), 1210–1219. <https://doi.org/10.1093/jnci/djae056> (2024).
11. Zehra, S., Doyle, F. & O'Malley, G. Letter to the Editor: Re: Association of risk factors and breast cancer among women treated at Tikur Anbessa Specialized Hospital. *BMJ Open*. **13** (2), e060636. <https://doi.org/10.1136/bmjopen-2022-060636> (2023).
12. Rweyemamu, L. P. et al. The distribution of reproductive risk factors disclosed the heterogeneity of receptor-defined breast cancer subtypes among Tanzanian women. *BMC Womens Health*. **21**, 423. <https://doi.org/10.1186/s12905-021-01536-6> (2021).
13. Mekonnen, D. et al. Advanced-stage presentation of breast cancer in Ethiopia: a cross-sectional study. *BMC Cancer*. **21** (1), 980. <https://doi.org/10.1186/s12885-021-08723-0> (2021).
14. Hammond, M. E. et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J. Clin. Oncol.* **28** (16), 2784–2795. <https://doi.org/10.1200/JCO.2009.25.6529> (2010).
15. Wolff, A. C. et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* **31** (31), 3997–4013. <https://doi.org/10.1200/JCO.2013.50.9984> (2013).
16. Allemani, C. et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* **391** (10125), 1023–1075. [https://doi.org/10.1016/S0140-6736\(17\)33326-3](https://doi.org/10.1016/S0140-6736(17)33326-3) (2018).
17. Jedy-Agba, E. et al. Cancer incidence in Nigeria: a report from population-based cancer registries. *Cancer Epidemiol.* **42**, 85–92. <https://doi.org/10.1016/j.canep.2016.03.009> (2016).
18. Ma, H. et al. Oral contraceptive use and breast cancer risk in African American women. *Cancer Causes Control*. **28** (6), 583–594. <https://doi.org/10.1007/s10552-017-0896-3> (2017).
19. Brewer, H. R., Jones, M. E., Schoemaker, M. J., Ashworth, A. & Swerdlow, A. J. Family history and risk of breast cancer: an analysis accounting for family structure. *Breast Cancer Res. Treat.* **165** (1), 193–200. <https://doi.org/10.1007/s10549-017-4325-6> (2017).
20. American Cancer Society. *Breast Cancer Facts & Figs. 2022–2024* (American Cancer Society, 2022).
21. Huo, D. et al. Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. *J. Clin. Oncol.* **27** (27), 4515–4521. <https://doi.org/10.1200/JCO.2008.19.4729> (2009).

22. Pagani, L. et al. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91** (1), 83–96. <https://doi.org/10.1016/j.ajhg.2012.05.015> (2012).
23. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol.* **13** (11), 1141–1151. [https://doi.org/10.1016/S1470-2045\(12\)70425-4](https://doi.org/10.1016/S1470-2045(12)70425-4) (2012).
24. Lambe, M. et al. Transient increase in the risk of breast cancer after giving birth. *N Engl. J. Med.* **331** (1), 5–9. <https://doi.org/10.1056/NEJM199407073310102> (1994).
25. Islami, F. et al. Breastfeeding and breast cancer risk by receptor status-a systematic review and meta-analysis. *Ann. Oncol.* **26** (12), 2398–2407 (2015).
26. Bernier, M. O., Plu-Bureau, G., Bossard, N., Ayzac, L. & Thalabard, J. C. Breastfeeding and risk of breast cancer: a meta-analysis of published studies. *Hum. Reprod. Update.* **6** (4), 374–386. <https://doi.org/10.1093/humupd/6.4.374> (2000).
27. Russo, J. & Russo, I. H. The role of pregnancy and breastfeeding in breast cancer etiology. *Horm. Mol. Biol. Clin. Investig.* **41** (3), 20200017. <https://doi.org/10.1515/hmbci-2020-0017> (2020).
28. Russo, J. & Russo, I. H. Biological basis of pregnancy and breastfeeding protection against breast cancer. *J. Mammary Gland Biol. Neoplasia.* **25** (3), 179–188. <https://doi.org/10.1007/s10911-020-09457-w> (2020).
29. Lipworth, L., Bailey, L. R. & Trichopoulos, D. History of breast-feeding in relation to breast cancer risk: a review of the epidemiologic literature. *J. Natl. Cancer Inst.* **92** (4), 302–312. <https://doi.org/10.1093/jnci/92.4.302> (2000).
30. Centers for Disease Control and Prevention. *Breastfeeding Report Card: United States, 2020* (CDC, 2020).
31. Anderson, W. F., Pfeiffer, R. M., Dores, G. M. & Sherman, M. E. Comparison of age distribution patterns for different breast cancer subtypes. *Cancer Epidemiol. Biomarkers Prev.* **23** (12), 2670–2678. <https://doi.org/10.1158/1055-9965.EPI-14-0585> (2014).

Figures

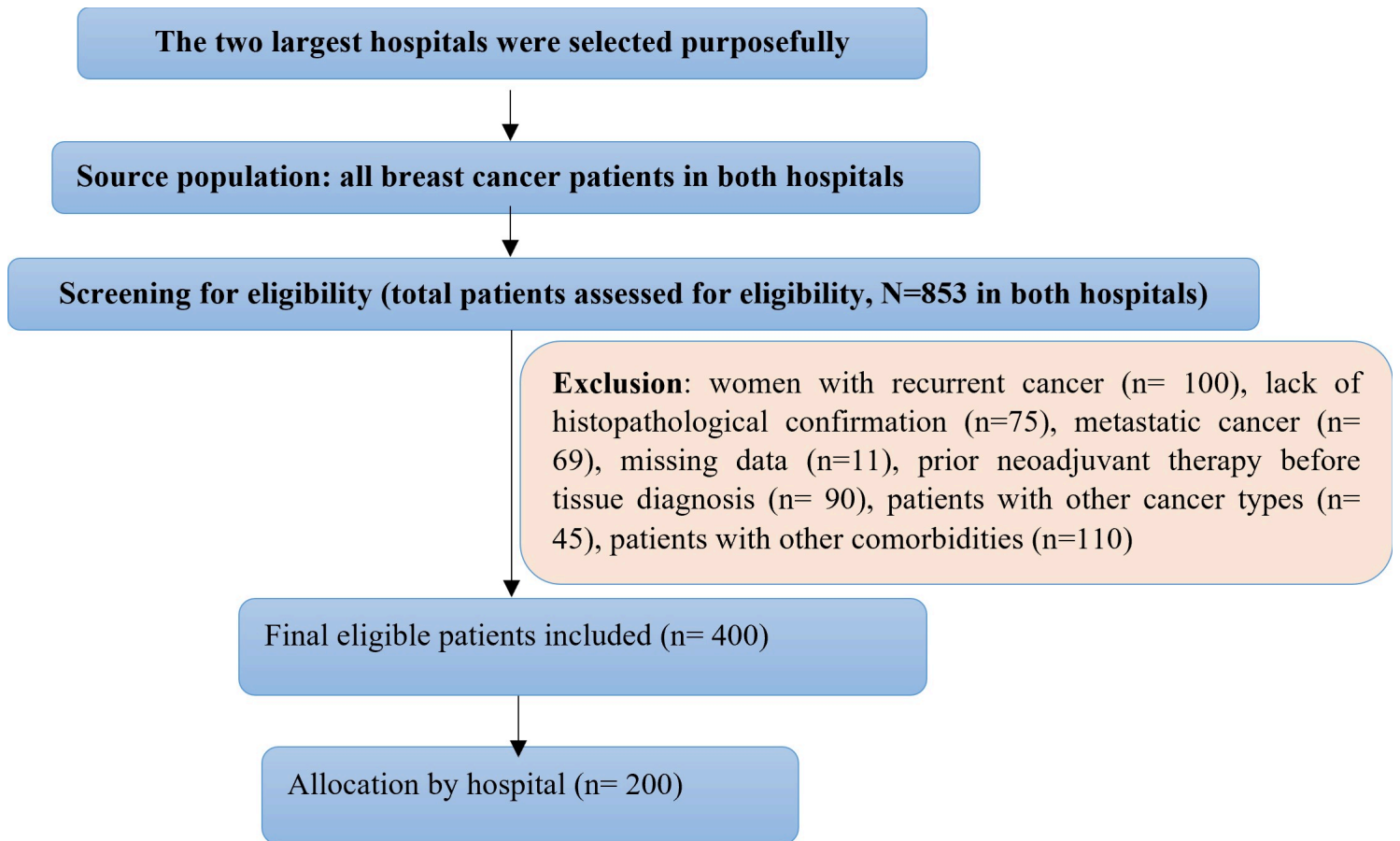


Figure 1
Sampling Flow Diagram

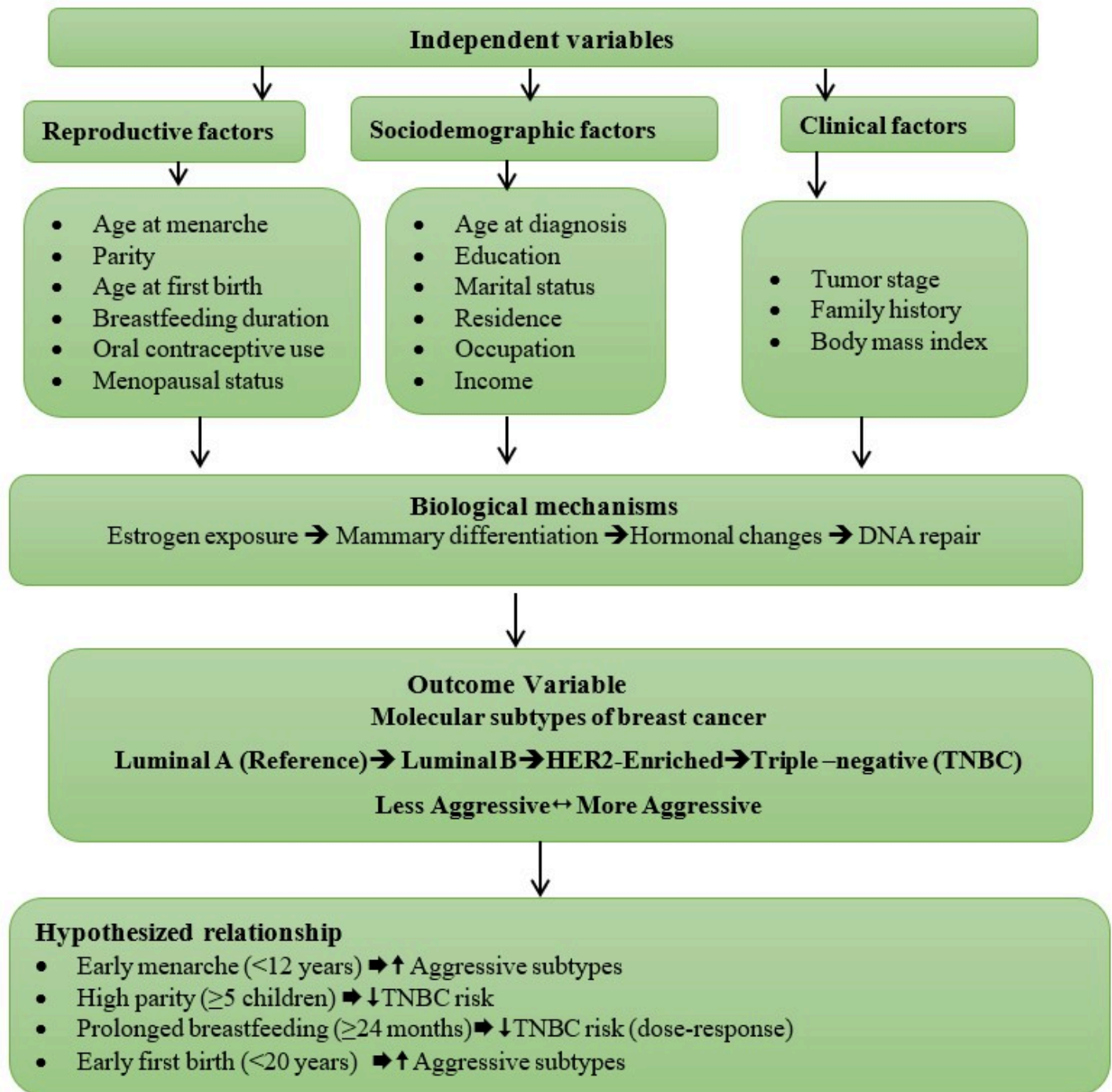


Figure 2

Conceptual framework illustrating the hypothesized relationship between multiple independent factors and aggressive breast cancer subtypes