Research Square

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

Spatio-temporal Analysis of Breast Cancer Incidence and Its Associated Risk Factors in Nigeria, 2021–2025: An Ecological Study Using Routine Health Information System Data

Yetunde Mariam Badmos

badmosyetunde@gmail.com

University of Ibadan

Olalekan John Taiwo

University of Ibadan

Odunayo Samuel Akomolede

University of Ibadan

Adewale Oluwagbenga Adeyefa

University of Ibadan

Research Article

Keywords: breast cancer, spatio-temporal analysis, spatial autocorrelation, geographically weighted regression, artificial light at night, DHIS2, Nigeria

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11511285/v1

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

Additional Declarations: No competing interests reported.

SPATIO-TEMPORAL ANALYSIS OF BREAST CANCER INCIDENCE AND ITS ASSOCIATED RISK FACTORS IN NIGERIA, 2021–2025: AN ECOLOGICAL STUDY USING ROUTINE HEALTH INFORMATION SYSTEM DATA

Yetunde Mariam Badmos,* Olalekan John Taiwo, Odunayo Samuel Akomolede, Adewale Oluwagbenga Adeyefa

Department of Geography, University of Ibadan, Ibadan, Nigeria

*Corresponding Author's Email and Phone Number: badmosyetunde@gmail.com, 07015860297

Abstract

Background: Reported breast cancer incidence in Nigeria is shaped by differences in diagnostic capacity and reporting between states. This examined the spatial and temporal pattern of DHIS2-reported breast cancer incidence across Nigerian states and its association with behavioural, socioeconomic and environmental factors.

Methods: In this ecological study, annual breast cancer case counts for the 36 states and the Federal Capital Territory (FCT) from 2021 to 2025 were obtained from the District Health Information Software 2 (DHIS2) and converted to incidence rates per 100,000 women. Trends were assessed with the Mann-Kendall trend test and Sen's slope, and spatial patterns with global and local Moran's I. Associations between 2025 incidence and 15 risk-factor variables were examined with Spearman correlation and bivariate Moran's I. Principal component analysis (PCA) reduced the variables to four components, used in ordinary least squares (OLS) and geographically weighted regression (GWR).

Results: Incidence varied widely, with extreme values in Sokoto (255.4 per 100,000) and Gombe (69.1) in 2022 that were probably reporting errors. No state showed a significant trend (national Sen's slope 0.13 per 100,000 per year; $p = 1.00$). Global Moran's I was not significant in any year (−0.055 to 0.058; $p = 0.39$–0.97). Local Moran's I identified only isolated clusters and outliers. Only formaldehyde ($r_s = -0.38$) and vegetation (NDVI; $r_s = -0.40$) correlated significantly with 2025 incidence, both negatively. Four components explained 82.7% of variance. In the OLS model ($R^2 = 0.264$; adjusted $R^2 = 0.172$), only the artificial light at night (ALAN) component was significant (coefficient 3.04; $p = 0.012$). GWR fitted better (AICc 250.9 versus 257.1; $R^2 = 0.344$), with local ALAN coefficients of 1.60 to 6.50, largest in the north-west.

Conclusions: Reported incidence showed no stable spatial clustering or temporal trend and probably reflects diagnostic access and reporting as much as disease occurrence. The spatially varying association with ALAN is ecological and exploratory. Wider cancer registration, including private facilities, is needed before spatial risk patterns can be interpreted confidently.

Keywords: breast cancer; spatio-temporal analysis; spatial autocorrelation; geographically weighted regression; artificial light at night; DHIS2; Nigeria

Background

Breast cancer is the most commonly diagnosed cancer among women worldwide. About 2.3 million women were diagnosed and about 666,000 died of the disease in 2022 [1]. In Africa, an estimated 198,300 new cases and 91,300 deaths occurred in 2022, and Nigeria recorded the largest numbers on the continent, with 32,278 new cases and 16,332 deaths [2, 3]. Reported incidence in sub-Saharan Africa is lower than in high-income countries, but mortality is disproportionately high. The difference is partly explained by population structure, risk-factor exposure, access to screening and diagnosis, stage at presentation, treatment availability and the completeness of cancer surveillance [4]. Population-based cancer registration in the region remains incomplete, and available evidence does not adequately represent the true burden [4]. In Nigeria, inadequate diagnostic infrastructure, financial barriers and delayed diagnosis continue to limit cancer control [5].

Breast cancer risk is shaped by non-modifiable, reproductive, behavioural, socioeconomic and environmental factors. An umbrella review found evidence linking obesity, alcohol consumption and smoking with higher risk and physical activity with lower risk [6], and a synthesis of systematic reviews reached similar conclusions about modifiable risk factors [7]. Obesity is associated with breast cancer, although the size and direction of the association vary with tumour characteristics and menopausal status [8], and a dose-response meta-analysis of more than seven million women reported positive associations for central obesity [9]. Alcohol consumption is associated with higher risk, including at low levels of intake [10]. Hormonal contraceptive use has been associated with a modest increase in risk, with the magnitude varying across studies [11] and with duration of use [12]. Socioeconomic conditions influence both exposure and detection. Poverty and educational attainment vary markedly between Nigerian states [13], and where diagnostic services are limited a low recorded incidence does not necessarily mean a low disease burden.

Environmental exposures may also vary geographically. A meta-analysis of 22 studies reported positive associations of long-term exposure to nitrogen dioxide (NO_2) and fine particulate matter ($PM_{2.5}$) with breast cancer incidence, with pooled hazard ratios of 1.02 per 10 $\mu g/m^3$ of NO_2 and 1.06 per 5 $\mu g/m^3$ of $PM_{2.5}$ [14], in line with earlier cohort evidence [15] and studies of specific toxic air pollutants [16]. Higher exposure to artificial light at night (ALAN) has been associated with increased risk, with moderate-quality evidence [17], particularly among premenopausal women and women with oestrogen-receptor-positive disease, although exposure measurement remains a limitation [18]. Residential greenness, measured by the Normalized Difference Vegetation Index (NDVI), was associated with lower breast cancer incidence in a large cohort in Israel [19]. These findings come from different settings and designs, and ecological associations cannot be read as individual-level causation, particularly in Nigeria, where urbanisation, industrial activity, income and health-service availability are strongly correlated.

Interpretation of spatial variation in Nigeria is further complicated by the quality of routine data. Facility-based reporting through DHIS2 can provide information across time and administrative areas, but studies in Gombe State [20] and in Ondo and Ekiti States [21] found variation in completeness, accuracy and consistency. Differences in reported breast cancer

incidence between states may therefore reflect disease occurrence, healthcare access, diagnostic capacity, facility distribution and reporting practice together.

National estimates show the scale of the burden but not how incidence varies between states over time, or whether states with high or low reported rates form geographical clusters. Studies of single risk factors also do not describe the combined distribution of behavioural, socioeconomic and environmental exposures, which are correlated with one another. Geographic information systems allow incidence to be mapped with these exposures. Spatial autocorrelation statistics test whether neighbouring areas have similar values and identify local clusters and outliers, principal component analysis (PCA) addresses correlation among predictors, and geographically weighted regression (GWR) allows relationships to vary across space.

To our knowledge, no previous study has examined reported breast cancer incidence across all Nigerian states over consecutive years using spatial statistics, or related it to behavioural, socioeconomic and remotely sensed environmental variables together. Existing Nigerian incidence estimates come from a small number of population-based registries [22, 23] or from modelled national figures [1], and neither describes how incidence varies between states over time. Geospatial analyses of breast cancer have been reported elsewhere [24]. Whether routine facility-reported data in Nigeria can support such an analysis has not, to our knowledge, been tested.

This study analysed reported breast cancer incidence across Nigeria from 2021 to 2025, using annual DHIS2 case counts to calculate state-level rates. The rates were examined alongside alcohol consumption, obesity, tobacco smoking, hormonal contraceptive use, educational attainment, poverty, income, ALAN, atmospheric pollutants, $PM_{2.5}$, land surface temperature and vegetation. Reported incidence was treated as a product of both disease occurrence and the capacity to detect and report cases. The aim was to describe the spatio-temporal distribution of reported breast cancer incidence and its associated behavioural, socioeconomic and environmental factors using GIS-based spatial statistical and regression techniques. The specific objectives were to describe the spatial distribution and temporal trends of incidence, identify global and local spatial clusters and outliers, examine bivariate relationships between incidence and selected risk factors, and assess global and spatially varying relationships between PCA-derived risk-factor dimensions and incidence using OLS regression and GWR.

Methods

Study setting

Nigeria lies on the Gulf of Guinea in West Africa and covers about 923,770 km² [25]. It comprises 36 states and the Federal Capital Territory (FCT), Abuja, which are grouped into six geopolitical zones: North Central, North East, North West, South East, South South and South West [25, 26]. The states are divided into 774 local government areas [25]. Nigeria is the most populous country in Africa, with an estimated 232.7 million people in 2024 [27]. The 36 states and the FCT were the 37 units of analysis in this study (Fig. 1).

The climate is tropical, with wet and dry seasons that vary by location. The rainy season is longest in the south, from March to November, and shortest in the far north, from mid-May to

September. Annual rainfall ranges from more than 3,000 mm in the south-east to no more than 500 mm in the far north [28]. Vegetation follows the same gradient, with mangrove, swamp and rainforest in the south and savanna grading from Guinea to Sudan to Sahel towards the north [28]. The harmattan, a hot, dry north-easterly wind, lasts more than three months in the north and rarely more than two weeks on the coast [28]. These gradients affect the remotely sensed variables used here, including land surface temperature and vegetation, and they run alongside differences in urbanisation, income, education and access to health care between the south and the north.

Health services are delivered at primary, secondary and tertiary levels. Local government authorities, with state governments, supervise primary care, state governments are mainly responsible for secondary care and the federal government for tertiary care. Private providers deliver an estimated 70% of services but operate only 35% of facilities [29]. Facilities report routine data to the national DHIS2 platform, which is the source of the case counts analysed here. No population-based cancer registry covers the whole country. As of 2015, six population-based registries operated in Abuja, Calabar, Ibadan, Enugu, Sokoto and Ekiti, and their completeness and data quality still needed improvement [30]. Cancer data from routine reporting may therefore differ in coverage and quality between states.



Fig. 1. *Study area map of Nigeria showing the 36 states, the Federal Capital Territory (FCT), the coast and the Niger and Benue rivers*

Source: Author's Analysis (2026)

Study design

This was an ecological, spatio-temporal study of Nigeria, which comprises 36 states and the FCT (37 units of analysis). Temporal and spatial analyses of incidence covered 2021 to 2025. Associations with risk factors were examined using the 2025 incidence rate (the most recent year available). The study was guided by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline.

Data sources

The datasets are summarised in Table 1. The outcome was the annual number of breast cancer cases reported through DHIS2 for each state and the FCT. DHIS2 is a facility-based reporting platform, so the resulting rates describe cases captured by the health information system and not those of a population-based cancer registry. Cases are assigned to the state of the reporting facility and not the patient's state of residence [31].

Behavioural and reproductive variables were taken from the 2024 Nigeria Demographic and Health Survey (NDHS) [26]. They were the proportion of women who consume alcohol, female obesity prevalence, female tobacco-smoking prevalence, hormonal contraceptive use and the proportion of women with no formal education. All were available by state and expressed as percentages. Household reliance on solid fuels for cooking was also taken from the NDHS and expressed as the percentage of households. Poverty was taken from the 2022 Nigeria Multidimensional Poverty Index [13], which provides subnational measures for all states and the FCT, and state-level income in purchasing-power-parity terms from the Global Data Lab Subnational Human Development Database [32].

Environmental variables were derived from remotely sensed and modelled products. ALAN was represented by annual mean night-time radiance from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band at about 500 m resolution ($nW/cm^2/sr$). Tropospheric formaldehyde (HCHO), sulphur dioxide (SO_2) and NO_2 were obtained from Sentinel-5P TROPOMI Level-2 products at about 7 km × 3.5 km ($\mu mol/m^2$) [33]. $PM_{2.5}$ ($\mu g/m^3$) came from the Copernicus Atmosphere Monitoring Service at about 5 km resolution [34]. Land surface temperature (LST, °C) was derived from MODIS products at about 1 km [35]. NDVI was derived from Landsat 8 Collection 2 Level-2 imagery at 30 m [19]. Annual composites were derived for 2021–2025.

Female population was obtained from WorldPop gridded estimates [36] at about 100 m resolution and summed by state, and was used as the denominator for the incidence rates. Administrative boundaries came from the FAO Global Administrative Unit Layers (GAUL) [37]. The datasets had different spatial resolutions and were harmonised on ArcGIS Pro to a common 1 km resolution.

Table 1. *Data sources and variables used in the analysis*

Variable	Source	Resolution or unit	Reference period
Breast cancer cases	DHIS2	State; annual counts	2021–2025
Female population	WorldPop	~100 m grid; persons	2021–2025

Variable	Source	Resolution or unit	Reference period
Alcohol consumption, obesity, tobacco smoking, hormonal contraceptive use, no formal education (women)	NDHS 2024	State; %	2024
Solid-fuel reliance	NDHS 2024	State; %	2024
Poverty	NBS Multidimensional Poverty Index	State; %	2022
Income	Global Data Lab Subnational Human Development Database	State; PPP	2022
ALAN	VIIRS Day/Night Band	~500 m; nW/cm^2/sr	2021–2025
HCHO, NO_2, SO_2	Sentinel-5P TROPOMI Level-2	~7 km × 3.5 km; µmol/m^2	2021–2025
$PM_{2.5}$	Copernicus Atmosphere Monitoring Service	~5 km; µg/m^3	2021–2025
Land surface temperature	MODIS	~1 km; °C	2021–2025
NDVI	Landsat 8 Collection 2 Level-2	30 m; unitless (−1 to +1)	2021–2025
Administrative boundaries	FAO GAUL	Vector	

Statistical analysis

Incidence rate

Annual state-level incidence rates were calculated as

$$IR = \frac{C}{P} \times 100{,}000$$

where *IR* is the incidence rate per 100,000 women, *C* the number of reported cases and *P* the female population. Rates are crude (not age-standardised).

Temporal trend

Trends in annual incidence rates were tested for each state, and for Nigeria as a whole, with the Mann–Kendall test, a non-parametric test that does not require normally distributed data and is less sensitive to outliers. The test statistic is

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}\left(x_j - x_k\right), \qquad \operatorname{sgn}\left(x_j - x_k\right) = \begin{cases} 1 & x_j > x_k \\ 0 & x_j = x_k \\ -1 & x_j < x_k \end{cases}$$

where x_j and x_k are observations at times j and k and n is the number of observations. Positive values of S indicate an increasing trend and negative values a decreasing trend. The standardised Z statistic was obtained from the variance of S, and significance was set at $p \leq 0.05$. The magnitude of the trend was estimated with Sen's slope, the median of all pairwise slopes:

$$Q_i = \frac{x_j - x_k}{j - k}, \qquad Q_{med} = \operatorname{median}(Q_i)$$

Spatial Autocorrelation

Global Moran's I was used to test whether incidence rates in neighbouring states were more similar (positive autocorrelation), less similar (negative) or unrelated (values near zero) than expected by chance [38]:

$$I = \frac{n}{W} \cdot \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where n is the number of units, w_{ij} the spatial weight between states i and j, W the sum of all weights, x_i the incidence rate in state i and $\bar{x}$ the mean rate. Local Indicators of Spatial Association (LISA) were then computed to identify high-high (HH) and low-low (LL) clusters and high-low (HL) and low-high (LH) spatial outliers:

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_j w_{ij} (x_j - \bar{x})$$

where S^2 is the variance of the variable. Spatial weights were based on Queen contiguity and row-standardised. Both statistics were computed for each year from 2021 to 2025 in ArcGIS Pro. Significance of global Moran's I was assessed with z-scores, and local significance with permutation-based p-values, at $p < 0.05$.

Bivariate analysis

The distribution of the 2025 incidence rate was tested with the Shapiro–Wilk test. Because it departed from normality and the number of units was small (n = 37), Spearman rank correlation was used to examine associations between incidence and each of the 15 explanatory variables and among the explanatory variables [39]. Correlations were interpreted as ecological associations at state level. Bivariate Moran's I, calculated in GeoDa with 999 permutations, was used to test whether incidence in a state was spatially associated with values of an explanatory variable in neighbouring states:

$$I_B = \frac{n}{S_0} \cdot \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{j=1}^{n} (y_j - \bar{y})^2}}$$

where x_i is the incidence rate in state i, y_j the value of the explanatory variable in neighbouring state j, $\bar{x}$ and $\bar{y}$ their means, and S_0 the sum of all spatial weights.

Principal component analysis

PCA was run in IBM SPSS Statistics on 15 variables: obesity, tobacco smoking, alcohol consumption, hormonal contraceptive use, solid-fuel reliance, land surface temperature, ALAN, HCHO, $PM_{2.5}$, NO_2, SO_2, poverty, no educational attainment, income and NDVI. Variables were selected because they have established or plausible associations with breast cancer risk or detection (see Background) and were available for all states. The correlation matrix was used so that variables were standardised. Suitability was checked with the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity. Components with eigenvalues above 1 were retained (Kaiser criterion) and rotated with Varimax rotation with Kaiser normalisation. Component scores were saved using the regression method and used as

explanatory variables in the regression models. Because components are orthogonal, this also removes collinearity among the predictors.

Regression analysis

OLS regression was run in ArcGIS Pro with the 2025 incidence rate as the dependent variable and the four component scores as explanatory variables:

$$Y = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \beta_4 F_4 + \varepsilon$$

where Y is the incidence rate per 100,000 women, β_0 the intercept, β_1 to β_4 the coefficients and F_1 to F_4 the component scores. Model performance was assessed with R^2, adjusted R^2, the corrected Akaike Information Criterion (AICc), the variance inflation factor (VIF), the Jarque–Bera statistic for normality of residuals and the Koenker (Breusch–Pagan) statistic for heteroskedasticity and non-stationarity.

GWR was then run in ArcGIS Pro to test whether relationships varied across space:

$$Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k (u_i, v_i) X_{ik} + \varepsilon_i$$

where (u_i, v_i) are the coordinates of state i and $\beta_k(u_i, v_i)$ is the local coefficient for explanatory variable k. Observations nearer to a target location receive greater weight through a kernel function. The ALAN component score was the single explanatory variable, because it was the only significant component in the OLS model, and a fixed distance bandwidth was selected by minimising AICc. GWR output included local coefficients, local R^2, predicted values and standardised residuals. GWR and OLS were compared using adjusted R^2, AICc and the spatial distribution of residuals. Results from both models were interpreted as ecological associations and not as individual-level causal effects.

Results

Spatial distribution of reported incidence

Reported incidence varied widely between states and years (Fig. 2). The highest rates per 100,000 women were most often in Zamfara (27.8, 31.9, 27.7, 17.4 and 19.4 from 2021 to 2025), Jigawa (22.6 in 2021, 12.1 in 2022, then 15.8, 18.4 and 19.3), Sokoto (16.7 in 2021 and 16.2 in 2025, with 22.4 and 20.9 in between), the FCT (14.6, 24.3, 21.9 and 12.1 from 2021 to 2024) and Kogi (14.2 to 16.6, apart from 2023). Lagos rose from 14.4 in 2021 to 21.7 in 2022, fell to 18.8 in 2023, peaked at 36.4 in 2024 and was 29.0 in 2025. Delta rose from 8.6 in 2021 to 24.3 in 2025 (13.3, 9.9 and 22.4 in between). Katsina increased from 4.7 in 2022 to 19.6 in 2025, Oyo reached 19.6 in 2025, and Ogun (10.5) and Ondo (9.2) rose in 2024. More states had rates above 10 per 100,000 in 2025 than in 2021.

Several values were extreme. Sokoto recorded 255.4 in 2022, far outside the range of every other state-year, and Gombe recorded 69.1 in 2022 (9.8 in 2021 and 22.2 in 2023). Kogi recorded 40.6 in 2023, up from 14.3, and Adamawa rose from 3.1 to 16.7 in 2025. These values were retained in the dataset and maps (and in the trend analysis, because they could not be verified; the Mann–Kendall test is rank-based and less sensitive to extreme values).

The lowest rates were consistently in southern and north-eastern states, including Ekiti (0.4 in

2021), Osun (0.8 in 2021 and 1.0 in 2025), Imo (1.1 in 2023 and 1.5 in 2025), Borno (2.2 in 2021 and 0.8 in 2025), Cross River (1.5 in 2025) and Akwa Ibom (2.4 in 2025).

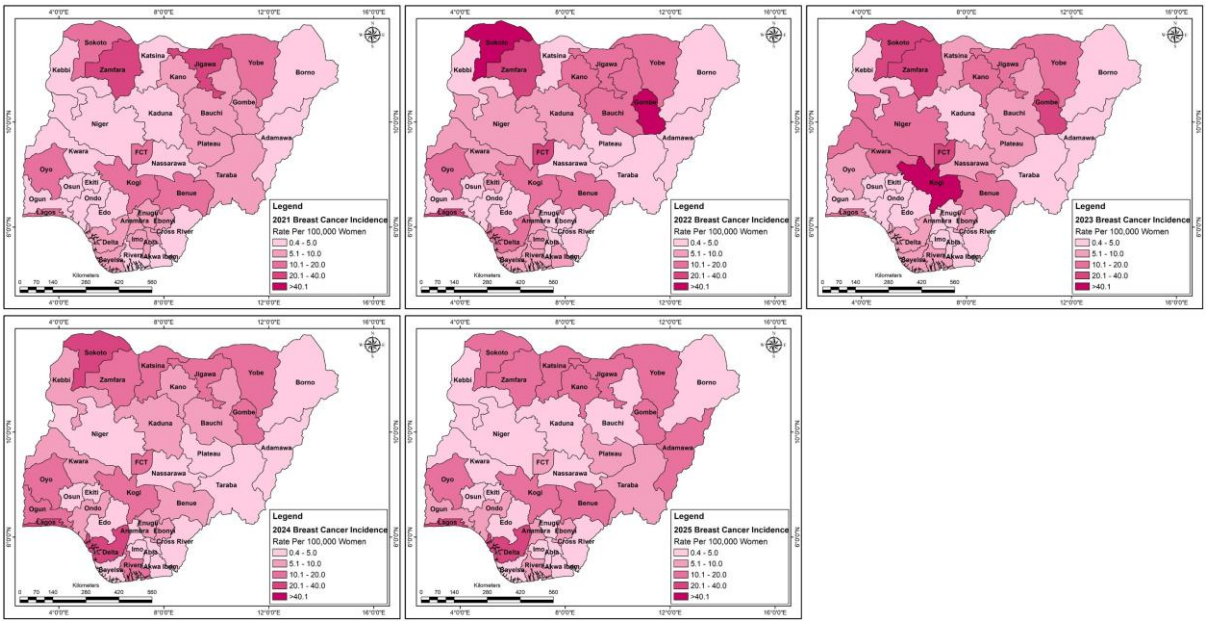


Fig. 2. *Spatial distribution of reported breast cancer incidence per 100,000 women in Nigeria, 2021 to 2025 (left to right, top to bottom)*

Source: Author's Analysis (2026)

Temporal trends

The Mann–Kendall test showed no significant national trend in incidence (Z = 0.000, p = 1.000), with a Sen's slope of 0.13 per 100,000 women per year (Table 2). No state had a significant trend. Sen's slope was positive in 18 of the 37 units and negative in 19. The steepest increases were in Lagos (+4.375 per 100,000 per year), Delta (+4.238) and Katsina (+4.100), and the steepest decreases in Zamfara (−3.808), the FCT (−3.000) and Imo (−2.088).

Table 2. *Mann–Kendall trend test and Sen's slope for breast cancer incidence rates, Nigeria and its states, 2021–2025. No trend was statistically significant (all p > 0.05)*

Level	State or territory	Sen's slope (per 100,000 per year)	Z	p	Trend direction
National	Nigeria	0.1257	0.0000	1.0000	Increasing
State	Abia	−0.9250	−1.0106	0.3122	Decreasing
State	Adamawa	1.6500	1.5159	0.1296	Increasing
State	Akwa Ibom	−0.2000	−0.7348	0.4624	Decreasing
State	Anambra	−0.7917	−0.7348	0.4624	Decreasing
State	Bauchi	−0.9750	−1.7146	0.0864	Decreasing
State	Bayelsa	−1.0417	−1.2247	0.2207	Decreasing
State	Benue	−1.4917	−1.2247	0.2207	Decreasing
State	Borno	−0.2333	−0.2449	0.8065	Decreasing
State	Cross River	−0.5000	−0.2449	0.8065	Decreasing
State	Delta	4.2375	1.7146	0.0864	Increasing
State	Ebonyi	0.7583	1.0106	0.3122	Increasing
State	Edo	−0.2875	−1.2247	0.2207	Decreasing
State	Ekiti	0.8750	0.7348	0.4624	Increasing
State	Enugu	−0.7000	−1.2247	0.2207	Decreasing
State	FCT	−3.0000	−1.2247	0.2207	Decreasing
State	Gombe	0.0667	0.0000	1.0000	Increasing
State	Imo	−2.0875	−0.7348	0.4624	Decreasing
State	Jigawa	1.3250	0.2449	0.8065	Increasing
State	Kaduna	0.2125	0.0000	1.0000	Increasing
State	Kano	1.0167	1.2247	0.2207	Increasing
State	Katsina	4.1000	1.7146	0.0864	Increasing
State	Kebbi	1.0500	1.7146	0.0864	Increasing
State	Kogi	−0.6000	−0.7348	0.4624	Decreasing
State	Kwara	0.1875	0.0000	1.0000	Increasing
State	Lagos	4.3750	1.2247	0.2207	Increasing
State	Nasarawa	−0.3292	−0.2449	0.8065	Decreasing
State	Niger	−0.2625	−0.5053	0.6134	Decreasing
State	Ogun	1.0500	1.0106	0.3122	Increasing
State	Ondo	1.1125	1.7146	0.0864	Increasing
State	Osun	0.0500	0.2449	0.8065	Increasing
State	Oyo	0.9000	0.2449	0.8065	Increasing
State	Plateau	−0.0375	0.0000	1.0000	Decreasing
State	Rivers	1.3000	1.2247	0.2207	Increasing
State	Sokoto	−2.3000	−0.7348	0.4624	Decreasing
State	Taraba	−0.1000	−0.2449	0.8065	Decreasing
State	Yobe	0.9250	0.7348	0.4624	Increasing
State	Zamfara	−3.8083	−1.2247	0.2207	Decreasing

Source: Author's Analysis (2026)

Global spatial autocorrelation

Global Moran's I was not significant in any year from 2021 to 2025, so incidence did not differ from a random spatial pattern (Table 3). Values ranged from −0.055 (2021) to 0.058 (2024), against an expected index of −0.028, and p-values ranged from 0.39 to 0.97.

Table 3. *Global Moran's I for breast cancer incidence rates, Nigerian states, 2021–2025*

Year	Moran's I	Expected I	Variance	z-score	p-value	Pattern
2021	−0.055278	−0.027778	0.010709	−0.265736	0.790443	Random
2022	0.008590	−0.027778	0.002024	0.808375	0.418875	Random
2023	0.013328	−0.027778	0.010153	0.407944	0.683315	Random
2024	0.058356	−0.027778	0.010136	0.855558	0.392242	Random
2025	−0.023975	−0.027778	0.011342	0.035706	0.971517	Random

Source: Author's Analysis (2026)

Local spatial autocorrelation

LISA identified a small number of significant clusters and outliers, none of which persisted across the period except Kebbi (Table 4; Fig. 3). In 2021, Katsina and Kebbi were LH outliers. In 2022, Kebbi was again an LH outlier (local I = −0.672; z = −3.655; p = 0.005). In 2023, Niger was an HH cluster (local I = 0.232; z = 2.852; p = 0.010), the only HH cluster in the study period, Kebbi was an LH outlier for a third consecutive year, and Abia and Rivers were LL clusters. In 2024, Delta was an HL outlier and Akwa Ibom and Ebonyi were LL clusters. In 2025, Abia was an LL cluster, Borno an LH outlier and Kogi an HL outlier.

Table 4. *Statistically significant local clusters and spatial outliers of breast cancer incidence (Local Moran's I, $p < 0.05$), 2021–2025. HH, high-high cluster; LL, low-low cluster; HL, high-low outlier; LH, low-high outlier*

Year	State	Local Moran's I	z-score	p-value	Type
2021	Katsina	−1.572467	−2.377498	0.021	LH
2021	Kebbi	−0.570089	−2.337778	0.021	LH
2022	Kebbi	−0.672184	−3.655497	0.005	LH
2023	Abia	0.377520	1.882968	0.007	LL
2023	Kebbi	−0.886329	−2.313346	0.029	LH
2023	Niger	0.231605	2.852334	0.010	HH
2023	Rivers	0.133552	1.583966	0.024	LL
2024	Akwa Ibom	0.423893	1.650328	0.025	LL
2024	Delta	−0.951864	−1.560098	0.044	HL
2024	Ebonyi	0.007555	1.656785	0.032	LL
2025	Abia	0.535048	2.007879	0.017	LL
2025	Borno	−1.274457	−1.999701	0.033	LH
2025	Kogi	−0.348735	−1.869840	0.014	HL

Source: Author's Analysis (2026)

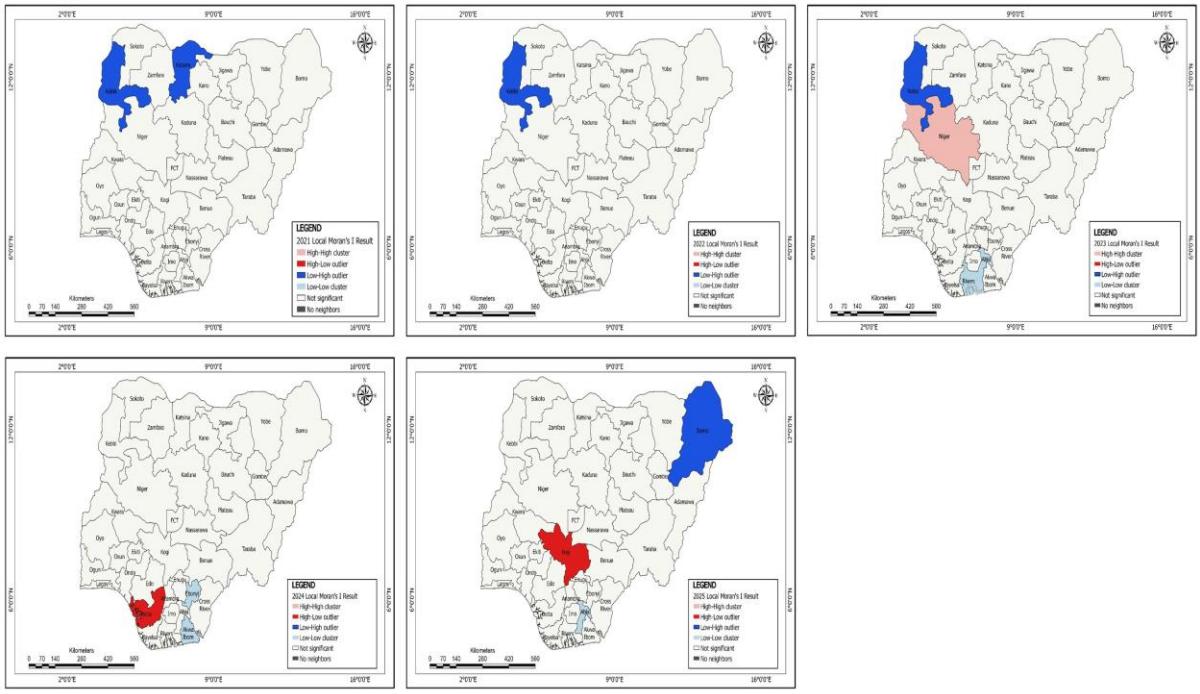


Fig. 3. *Local clusters and spatial outliers of breast cancer incidence in Nigeria (Local Moran's I), 2021 to 2025 (left to right, top to bottom)*

Source: Author's Analysis (2026)

Associations with risk factors in 2025

The 2025 incidence rate was right-skewed, with a long tail towards high values, and departed significantly from normality (Shapiro–Wilk W = 0.876, p = 0.0007; Fig. 4). Spearman correlation was therefore used (Fig. 5). Only two of the 15 variables were significantly correlated with incidence: HCHO ($r_s = -0.38$, $p < 0.05$) and NDVI ($r_s = -0.40$, $p < 0.05$), both negatively. ALAN, land surface temperature, NO_2, $PM_{2.5}$, SO_2, solid-fuel reliance, no educational attainment, poverty, income, obesity, tobacco smoking, alcohol consumption and hormonal contraceptive use were not significantly correlated with incidence ($p > 0.05$).

The explanatory variables were strongly intercorrelated. The strongest correlations were between HCHO and NO_2 ($r_s = 0.92$, $p < 0.01$), land surface temperature and NDVI ($r_s = -0.88$), solid-fuel reliance and no educational attainment ($r_s = 0.85$), solid-fuel reliance and obesity prevalence ($r_s = -0.86$), income and NO_2 ($r_s = 0.84$), and income and obesity prevalence ($r_s = 0.77$).

Bivariate Moran's I showed significant positive spatial association between incidence and $PM_{2.5}$, land surface temperature, poverty and no educational attainment, the latter two with the strongest associations (Table 5). Significant negative spatial associations were found for HCHO, NO_2, solid-fuel reliance, income, tobacco smoking and NDVI. Associations with SO_2, obesity, hormonal contraceptive use and alcohol consumption were not significant, and the association with ALAN had p = 0.050.

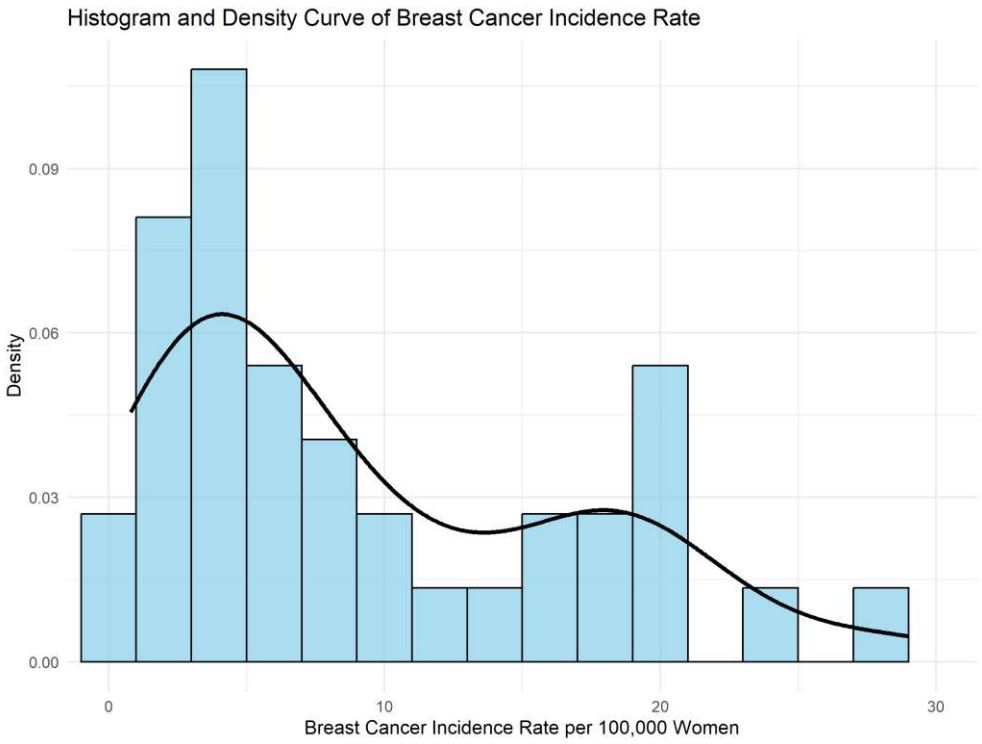


Fig. 4*. Histogram and density curve of the 2025 breast cancer incidence rate across the 37 units*

Source: Author's Analysis (2026)

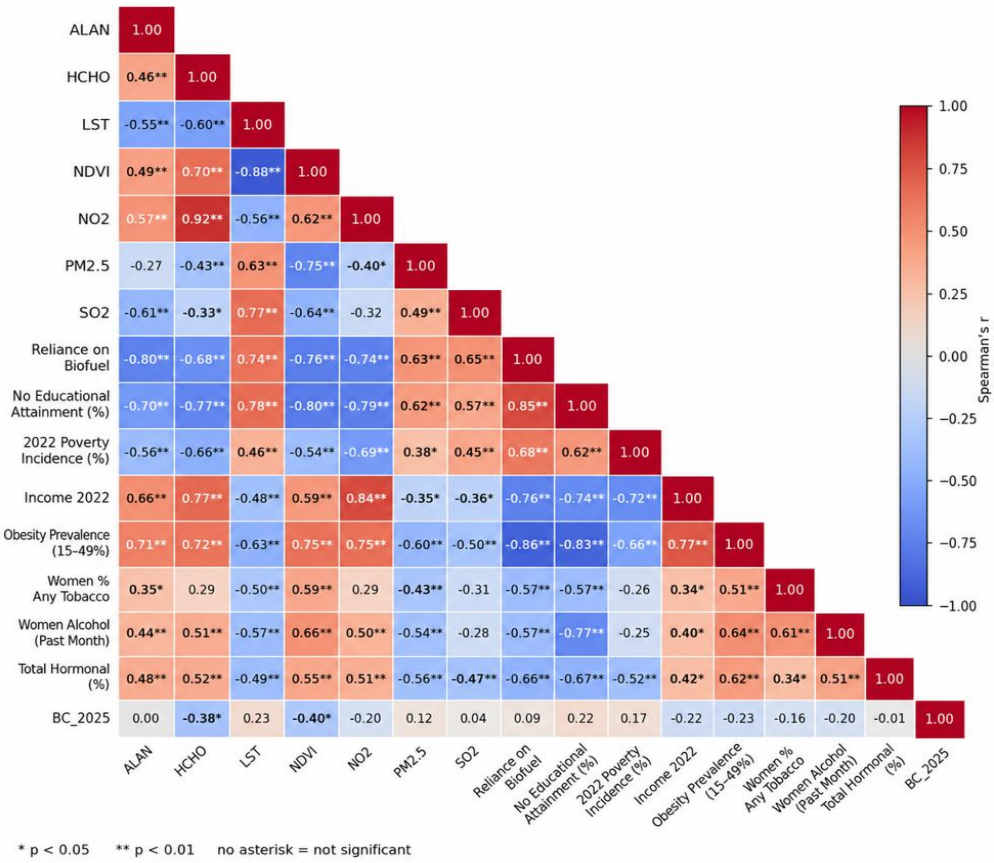


Fig. 5. *Spearman correlation matrix of breast cancer incidence and the 15 explanatory variables*

Source: Author's Analysis (2026)

Table 5. *Global bivariate Moran's I between the 2025 breast cancer incidence rate and risk factors*

Category	Variable	Moran's I	z-score	p-value	Significance
Environmental	ALAN	−0.118	−1.4369	0.050	ns
Environmental	HCHO	−0.280	−3.4301	0.002	**
Contextual	Land surface temperature	0.148	1.8997	0.038	*
Contextual	NDVI	−0.151	−1.8961	0.040	*
Environmental	NO_2	−0.145	−1.9410	0.025	*
Environmental	$PM_{2.5}$	0.143	1.8648	0.041	*
Environmental	SO_2	0.118	1.4709	0.064	ns
Environmental	Solid-fuel reliance	−0.130	−5.2156	0.001	**
Socioeconomic	No educational attainment	0.219	2.6579	0.004	**
Socioeconomic	Poverty (2022)	0.233	2.8507	0.006	**
Socioeconomic	Income (2022)	−0.209	−2.4883	0.009	**
Lifestyle and behavioural	Obesity prevalence	−0.136	−1.6602	0.055	ns
Lifestyle and behavioural	Tobacco smoking	−0.139	−1.7292	0.037	*
Lifestyle and behavioural	Hormonal contraceptive use	−0.105	−1.3072	0.083	ns
Lifestyle and behavioural	Alcohol consumption	−0.103	−1.2404	0.101	ns

Note: *** p < 0.01; * p < 0.05; ns, not significant*

Source: Author's Analysis (2026)

Principal component analysis

The data were suitable for PCA (KMO = 0.796; Bartlett's test $\chi^2 = 555.081$, df = 105, $p < 0.001$). Four components had eigenvalues above 1 and together explained 82.72% of the total variance (Table 6). After Varimax rotation they explained 30.15%, 21.75%, 20.35% and 10.47%.

Component 1 (Socioeconomic and Air Pollution) had positive loadings for NO_2 (0.888), income (0.865), HCHO (0.813) and obesity prevalence (0.687) and negative loadings for poverty (−0.714), no educational attainment (−0.615) and solid-fuel reliance (−0.540). Component 2 (Lifestyle and Ecological) was defined by alcohol consumption (0.872), tobacco smoking (0.845) and NDVI (0.624), with negative loadings for no educational attainment (−0.561), land surface temperature (−0.531) and $PM_{2.5}$ (−0.513). Component 3 (Thermal and Chemical Environmental) had positive loadings for SO_2 (0.870) and land surface temperature (0.664) and negative loadings for hormonal contraceptive use (−0.682) and NDVI (−0.596). Component 4 (Artificial Light at Night) was defined almost entirely by ALAN (0.871).

Table 6. *Rotated component matrix*

Variable	Component 1	Component 2	Component 3	Component 4
ALAN	0.314	0.220	−0.129	**0.871**
HCHO	**0.813**	0.352	−0.301	−0.129
Land surface temperature	−0.298	−0.531	**0.664**	−0.171
NDVI	0.382	**0.624**	−0.596	0.000
NO_2	**0.888**	0.109	−0.136	0.273
$PM_{2.5}$	−0.358	**−0.513**	0.395	0.124
SO_2	−0.057	−0.134	**0.870**	−0.394

Variable	Component 1	Component 2	Component 3	Component 4
Solid-fuel reliance	**−0.540**	−0.351	0.487	−0.441
No educational attainment	**−0.615**	−0.561	0.430	−0.098
Poverty	**−0.714**	0.032	0.486	−0.125
Income	**0.865**	0.164	−0.112	0.283
Obesity prevalence	**0.687**	0.418	−0.254	0.373
Tobacco smoking	0.008	**0.845**	−0.122	0.165
Alcohol consumption	0.201	**0.872**	−0.058	0.213
Hormonal contraceptive use	0.410	0.112	**−0.682**	−0.106

Note: *Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 10 iterations. Bold, primary loading of each variable*

Source: Author's Analysis (2026)

Ordinary least squares regression

The OLS model explained 26.4% of the variance in incidence ($R^2 = 0.264$; adjusted $R^2 = 0.172$), and the joint F-statistic was significant ($F = 2.876$, $p = 0.038$; Table 7). Only Component 4 (ALAN) was a significant predictor (coefficient 3.038, 95% CI 0.70 to 5.37, $p = 0.012$; $p = 0.046$ with heteroskedasticity-consistent standard errors). A one-unit increase in the ALAN component score was associated with about 3.04 additional cases per 100,000 women. Components 1 (−1.484, $p = 0.205$), 2 (−1.887, $p = 0.109$) and 3 (0.336, $p = 0.771$) were not significant. All VIF values were 1.000, as expected for orthogonal components. Residuals did not depart from normality (Jarque–Bera $p = 0.467$), and the Koenker statistic was not significant ($p = 0.182$). AICc was 257.105.

Table 7. *Ordinary least squares regression of the 2025 breast cancer incidence rate on the four component scores.*

Variable	Coefficient	Std. error	t	p	t (HC)	p (HC)	VIF
Intercept	9.278	1.130	8.208	<0.001	8.826	<0.001	n/a
Component 1	−1.484	1.146	−1.295	0.205	−1.518	0.139	1.000
Component 2	−1.887	1.146	−1.647	0.109	−1.493	0.145	1.000
Component 3	0.336	1.146	0.293	0.771	0.344	0.733	1.000
Component 4 (ALAN)	3.038	1.146	2.651	0.012	2.080	0.046	1.000

Note: *$R^2 = 0.264$; adjusted $R^2 = 0.172$; $F(4, 32) = 2.876$, $p = 0.038$; AICc = 257.105; Koenker (BP) = 6.238, df = 4, p = 0.182; Jarque-Bera = 1.524, p = 0.467. CI, confidence interval, calculated from the conventional standard errors (t distribution, 32 residual degrees of freedom); HC, heteroskedasticity-consistent standard errors; VIF, variance inflation factor.*

Source: Author's Analysis (2026)

Geographically weighted regression

The GWR model had a lower AICc than the OLS model (250.886 versus 257.105) and explained more variance ($R^2 = 0.344$; adjusted $R^2 = 0.247$; Table 8). Local R^2 was highest in Lagos and Oyo (0.266–0.318) and lowest in Borno, Yobe and Adamawa (0.042–0.086; Fig. 6). Local coefficients for the ALAN component ranged from 1.604 in the south to 6.501 in the north-west, with the largest values in Sokoto, Kebbi and Zamfara and the smallest in the

South-South and South-East (Fig. 7). All standardised residuals were within ±2.5 standard deviations (Fig. 8).

Table 8. *Geographically weighted regression diagnostics*

Statistic	Value
Bandwidth (as reported by the software)	466,925.66
Residual sum of squares	1,349.12
Effective number of parameters	5.64
Sigma	6.559
AICc	250.886
R^2	0.344
Adjusted R^2	0.247

Source: Author's Analysis (2026)

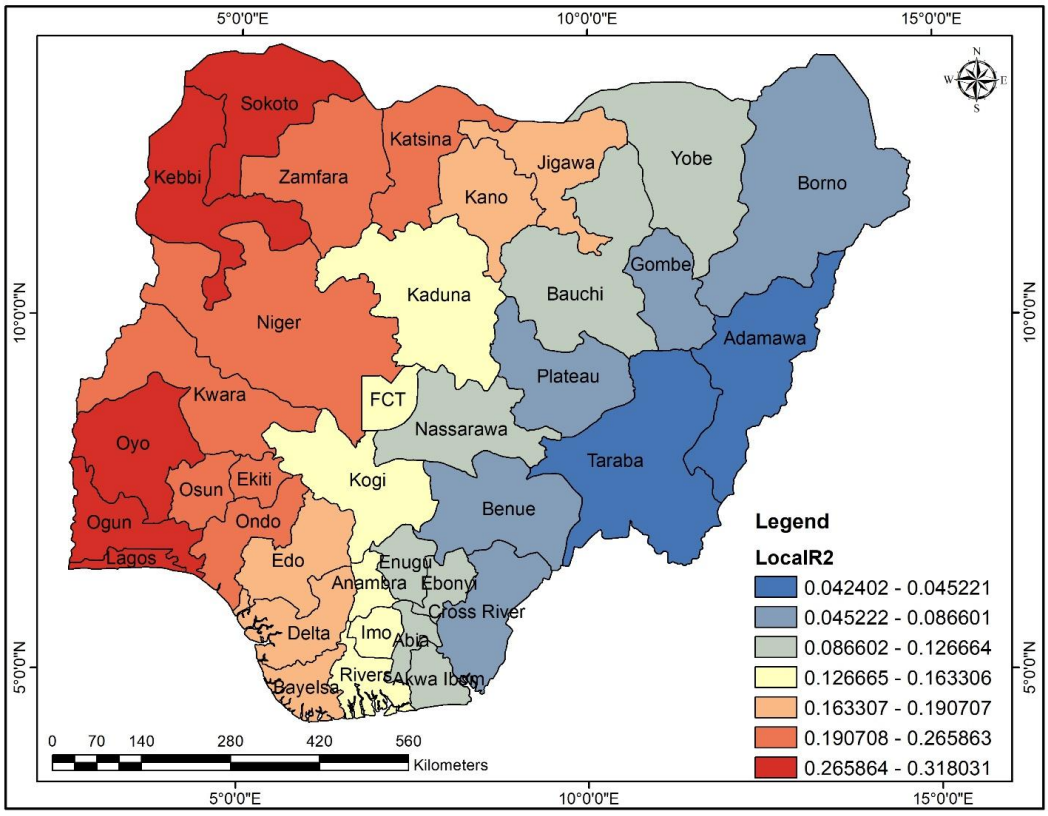


Fig. 6. *Local R^2 of the geographical weighted regression for the ALAN component*

Source: Author's Analysis (2026)

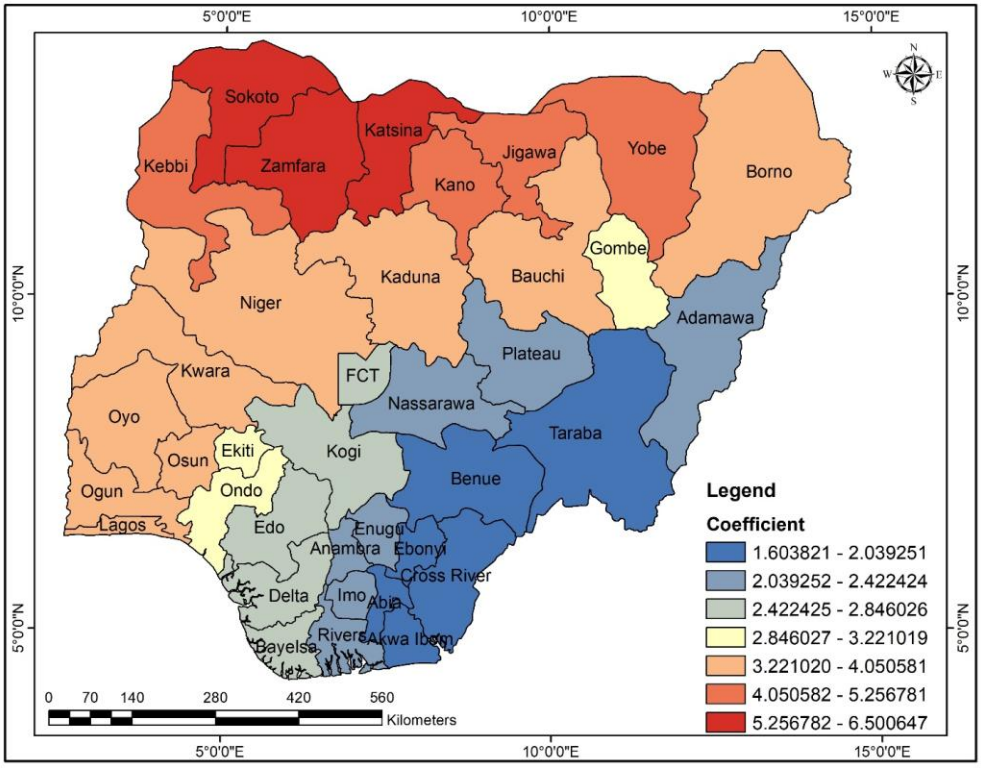


Fig. 7. *Local coefficients of the ALAN component in the geographically weighted regression*

Source: Author's Analysis (2026)

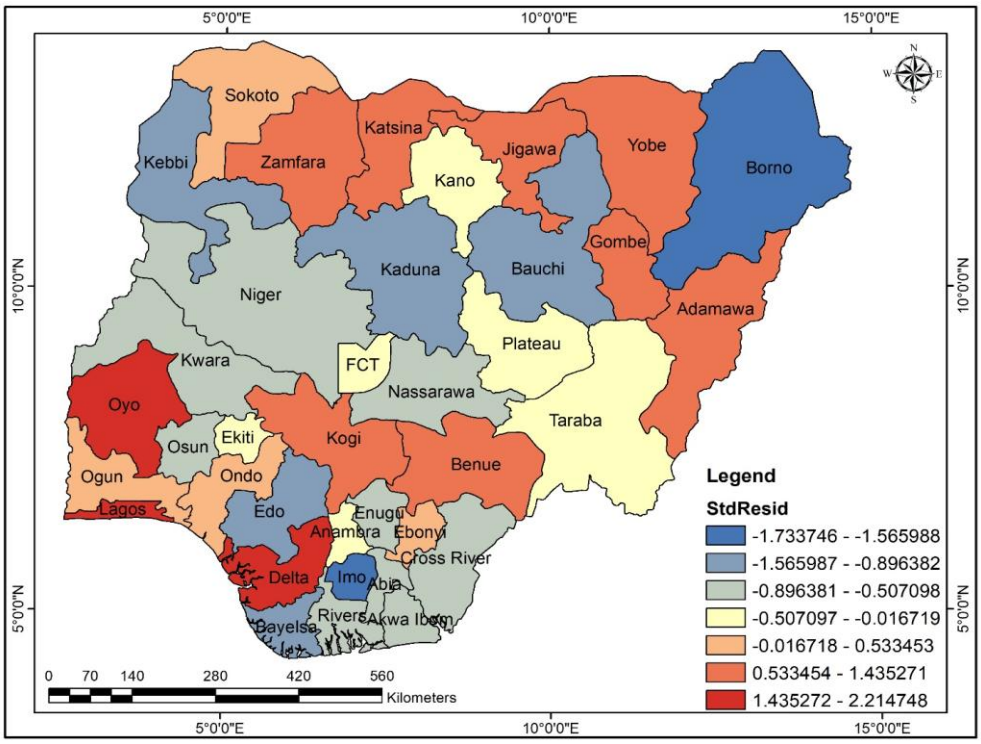


Fig. 8. *Standardised residuals of the geographically weighted regression for the ALAN component*

Source: Author's Analysis (2026)

Discussion

Principal findings

Reported breast cancer incidence in Nigeria between 2021 and 2025 showed no significant temporal trend in any state, no significant global spatial clustering in any year, and only a few isolated local clusters and outliers. Of the 15 behavioural, socioeconomic and environmental variables, only HCHO and NDVI correlated with 2025 incidence, both negatively. After reduction to four components, only a component dominated by ALAN was associated with incidence, and GWR indicated that the strength of this association differed across the country. The pattern is most plausibly explained by the combined influence of disease occurrence, access to diagnosis and the completeness of reporting, and the data cannot separate these.

This study adds two things to existing evidence. To our knowledge, it provides the first state-level description of DHIS2-reported breast cancer incidence in Nigeria across several years, together with survey-based and remotely sensed exposures. It also shows that these routine data, in their current form, do not yield a spatial or temporal signal that can be separated from reporting practice: incidence was spatially random in every year, the extreme values were implausible, and associations with known carcinogens ran opposite to expectation. This is informative for cancer surveillance planning because it identifies where routine data fall short.

Reported incidence and health-system capacity

Several features of the maps point to reporting and access as major determinants of recorded incidence. Zamfara, Jigawa, Sokoto and Yobe had high rates in 2021, yet these states are among the most socioeconomically deprived in Nigeria, with low female literacy and limited access to care [13]. Women in northern Nigeria often present to public hospitals only when disease is advanced [40], so the cases that are recorded may be late-stage and more visible in hospital records [41, 42]. High rates in Lagos and the FCT are expected, because urban populations reach public facilities more readily and urbanisation brings lifestyle changes associated with breast cancer, including physical inactivity, obesity, alcohol use, delayed childbearing and reduced breastfeeding [43, 44]. Higher breast cancer burden in Lagos has also been documented by cancer registry data [45]. The rise in Lagos to 36.4 per 100,000 in 2024 and the sustained increase in Delta are compatible with a growing role of such risk factors, but they are equally compatible with better detection and reporting [41, 45]. Oil-related pollution has been raised as a concern in the Niger Delta [46], but the present data cannot attribute Delta's trend to it.

Low rates in Ekiti, Osun, Kebbi, Imo, Akwa Ibom and Borno probably reflect limited public-sector oncology services, private-sector dominance in healthcare, low use of public facilities, and in Borno the destruction of health infrastructure by insecurity, and not a low underlying burden [22, 40, 42]. The same applies to Imo's fall to 1.1 per 100,000 in 2023.

The extreme single-year values support this interpretation. The 2022 rate of 255.4 per 100,000 in Sokoto exceeds the highest population-based rates generally reported worldwide [43, 45] and is probably a data entry or reporting error, such as several years of records entered in one reporting cycle or a duplicate entry. The jump in Gombe from 9.8 to 69.1 is

similarly implausible as a real change, and both values fell back to about 22 in 2023. Kogi's 2023 value of 40.6 may partly reflect referral patterns, as DHIS2 assigns cases to the facility's state and not the patient's [31]. The Adamawa rise from 3.1 to 16.7 in 2025 may reflect backlogged cases entered late in a state affected by disruption to health services [40]. The elevated rate in Oyo in 2025 may reflect improved documentation at the University College Hospital, Ibadan, a principal oncology referral centre [47]. Declines in the FCT and Benue between 2023 and 2024 cannot be explained from these data and may reflect reduced facility attendance or reporting gaps. Such unexplained year-to-year change is common in routine health data in sub-Saharan Africa [48, 49], and it argues for reading multi-year patterns in preference to single-year values. The increase in the number of states above 10 per 100,000 between 2021 and 2025 is consistent with gradual widening of DHIS2 reporting coverage and not necessarily with rising disease occurrence.

Temporal and spatial patterns

The absence of any significant trend should not be read as evidence of stability. With five annual observations per state, the Mann–Kendall test has little power, and the Sen's slope estimates, although small, show the direction of change. The steepest increases (Lagos, Delta, Katsina) and decreases (Zamfara, FCT, Imo) could each reflect changes in reporting as much as in incidence [48, 49].

A random global pattern in every year means that neighbouring states did not tend to share similar rates. This is consistent with findings elsewhere in sub-Saharan Africa, where regional cancer incidence is heterogeneous and partly shaped by health-system factors [1, 23]. DHIS2-reported incidence depends on the number of reporting facilities, the quality of documentation and whether patients from other states are counted, all of which are administratively determined and do not follow geographical contiguity [23, 50]. Breast cancer is also driven mainly by behavioural, hormonal and socioeconomic factors that do not spread between neighbouring states as environmentally determined diseases may [51, 52]. A non-significant global statistic does not exclude local clusters [53, 54], which is why LISA was applied.

The local results were sparse and unstable. Kebbi was an LH outlier for three consecutive years (2021–2023), a pattern that fits high solid-fuel use in a state with limited diagnostic infrastructure and low alcohol and hormonal contraceptive use [55- 57]. Niger was the only HH cluster (2023), in a zone with increasing urbanisation, ALAN, air pollution and obesity [58, 59]. Delta's HL classification in 2024 and Kogi's in 2025 could indicate local environmental exposures, such as gas flaring or mining [59, 60], but also local differences in case detection. Borno's LH classification fits poor diagnostic access and insecurity [61]. Because most classifications appeared in a single year, they are better read as hypotheses than as established clusters.

Risk factors

The negative correlation of HCHO with incidence ($r_s = -0.38$) contradicts the established carcinogenicity of formaldehyde [62]. It is most likely explained by underdiagnosis in the more polluted northern states, where access to mammography and diagnostic services is poor

[57]. This is consistent with evidence that reported rates in sub-Saharan Africa reflect health-system capacity [4] and that about 76% of Nigerian cases diagnosed between 2018 and 2023 were at an advanced stage [63]. The negative correlation with NDVI ($r_s = -0.40$) is in the expected direction, as greenspace exposure has been associated with lower incidence of breast and other cancers through effects on air pollution, physical activity, stress and living conditions [64], although in Nigeria greenness also varies with latitude, urbanisation and health-service coverage.

The bivariate Moran's I results had mixed directions. Poverty and low educational attainment, $PM_{2.5}$ and land surface temperature were positively associated with incidence in neighbouring states, whereas HCHO, NO_2, solid-fuel reliance, income, tobacco smoking and NDVI were negatively associated. The negative associations for known carcinogens again fit a geographical mismatch between where environmental and socioeconomic burdens are concentrated, mostly in the north, and where cases are reported [65, 66]. The positive associations with poverty and low education may indicate that deprived areas also have higher reported incidence, possibly through late-stage presentation, but that interpretation sits uneasily with the negative associations above and with the non-significant Spearman results for the same variables. Together, the inconsistent directions suggest that these associations are shaped by reporting patterns more than by risk.

PCA grouped the 15 variables into four dimensions that match known structure in Nigeria: a socioeconomic and air pollution gradient between wealthier urban states and poorer states reliant on solid fuels [57]; a lifestyle and ecological dimension in which alcohol and tobacco use co-occur with vegetation cover, mainly in southern states; a thermal and chemical dimension combining SO_2 and heat; and a separate ALAN dimension. Alcohol is classified by IARC as carcinogenic to the breast [67], and smoking and hormonal contraceptive use have been linked to increased risk [68], yet none of these variables was significantly associated with incidence in the present study. This is more likely to reflect the limits of state-level data on a small number of units than evidence against these associations.

The significant association between the ALAN component and incidence in OLS, and its spatial variation in GWR, agree in direction with epidemiological evidence linking outdoor ALAN with breast cancer, possibly through disruption of circadian rhythm and melatonin [59, 60, 69, 70]. There are two reasons for caution. First, night-time radiance also captures urbanisation, industrial activity and the density of health facilities, so an association with reported incidence may reflect better case detection in brighter states and not a carcinogenic effect of light. Second, the model had only 37 observations, the ALAN coefficient was less certain with heteroskedasticity-consistent standard errors ($p = 0.046$), and the model explained 26.4% of variance (adjusted 17.2%). The Koenker statistic was not significant, so the evidence of non-stationarity rests mainly on the lower AICc of the GWR model (a difference of 6.2) and should be treated as exploratory. The strongest local coefficients were in Sokoto, Kebbi and Zamfara, where reported incidence varied considerably between years, so the local estimates may be influenced by a few states. Earlier work suggests that the influence of ALAN differs between geographic contexts and may be strongest where other

protective factors are limited [60]. The weakest local fit in the north-east (local R^2 0.042–0.086) coincides with the area where surveillance is most disrupted.

Implications for surveillance and prevention

The main implication is that incidence derived from DHIS2 cannot yet be used alone to describe the geography of breast cancer in Nigeria. Three actions follow from the findings. First, national health information policy should require private hospitals, clinics and faith-based facilities to report cancer cases to DHIS2 or an integrated national cancer registry, because their absence from routine data limits planning and resource allocation. Second, state-level cancer registries should be expanded, especially in states with very low reported incidence such as Borno, Imo and Akwa Ibom, where low burden cannot be distinguished from under-reporting, and in the north-east, where insecurity disrupts facility operations and data collection. Third, breast cancer awareness and early detection programmes should be strengthened nationwide, with messaging adapted to regional and cultural context, because late presentation is common [63]. Environmental exposure monitoring, including air pollution and night-time lighting, could be added to public health planning, particularly where industrial exposure is concentrated, as in the Niger Delta. The present evidence does not show that these exposures increase risk in Nigeria.

Strengths and limitations

The study combines epidemiological, behavioural, socioeconomic and remotely sensed data for all 36 states and the FCT, applies global and local spatial statistics, and treats data quality explicitly in interpretation. It also has important limitations.

DHIS2 captures records mainly from public facilities, so private, faith-based and non-governmental facilities are not represented. This probably leads to under-estimation in southern states such as Imo, Osun and Akwa Ibom, where private care is common. Insecurity has disrupted routine reporting in Borno, Yobe and parts of Adamawa, so low rates there may reflect weak surveillance. Cases are assigned to the facility's state and not the patient's, rates were not age-standardised, and extreme values in 2022 and 2023 were retained. The five-year study period limits the power of trend analysis.

The analyses are ecological, with 37 units, and cannot support individual-level inference. Many correlations were tested without adjustment for multiple comparisons. A PCA of 15 variables on 37 observations is below commonly cited sample-size guidance, and the component scores, OLS and GWR models are sensitive to the few influential states. Risk-factor variables came from different years (for example 2022 and 2024) and were compared with 2025 incidence only. Datasets with different resolutions were harmonised to 1 km and summarised by state, which introduces spatial uncertainty and hides within-state variation in exposure, behaviour and incidence.

Future work should combine incidence with facility-level data on mammography, diagnostic capacity and screening services, including the private sector, to test whether mismatch between risk and reported incidence reflects access and reporting. Analysis at Local Government Area or facility level, where data allow, would show whether the randomness at

state level conceals local clustering. As DHIS2 records accumulate and private-sector reporting is integrated, longer time series will permit trend analysis with greater power. Dedicated studies are also needed in the conflict-affected north-east, where surveillance gaps are largest.

Conclusions

Reported breast cancer incidence in Nigeria between 2021 and 2025 showed no significant temporal trend in any state and no significant spatial clustering at the state level. Local clusters and outliers were few and mostly confined to single years. An ALAN-dominated component was the only significant predictor of incidence in OLS, and GWR indicated that this association varied across the country, but the finding is exploratory and ecological, and ALAN may stand in for urbanisation and health-service availability. Reported incidence does not fully capture the spatial burden of breast cancer in Nigeria. A clearer picture needs complete cancer registration that includes private facilities and conflict-affected states, alongside continued attention to risk-factor exposure.

Abbreviations

AICc: corrected Akaike Information Criterion; ALAN: artificial light at night; CAMS: Copernicus Atmosphere Monitoring Service; DHIS2: District Health Information Software 2; FCT: Federal Capital Territory; GAUL: Global Administrative Unit Layers; GWR: geographically weighted regression; HCHO: formaldehyde; HH: high-high; HL: high-low; KMO: Kaiser–Meyer–Olkin; LH: low-high; LISA: Local Indicators of Spatial Association; LL: low-low; LST: land surface temperature; MODIS: Moderate Resolution Imaging Spectroradiometer; NBS: National Bureau of Statistics; NDHS: Nigeria Demographic and Health Survey; NDVI: Normalized Difference Vegetation Index; NO_2: nitrogen dioxide; OLS: ordinary least squares; PCA: principal component analysis; $PM_{2.5}$: fine particulate matter; SO_2: sulphur dioxide; TROPOMI: TROPOspheric Monitoring Instrument; VIF: variance inflation factor; VIIRS: Visible Infrared Imaging Radiometer Suite.

Declarations

Ethics approval and consent to participate: Not applicable. This study used aggregated, de-identified secondary data and did not involve human participants.

Consent for publication: Not applicable.

Availability of data and materials: The DHS data analysed are available from the DHS Program (https://dhsprogram.com). The Multidimensional Poverty Index, Global Data Lab, VIIRS, Sentinel-5P, CAMS, MODIS, Landsat 8, WorldPop and FAO GAUL datasets are publicly available from the sources cited. The 2021-2025 breast cancer data analysed during this study is included in the supplementary information file.

Competing interests: The author declares that there are no competing interests.

Funding: No specific funding was received for this study.

Authors' contributions: YMB conceived and designed the study, acquired and analysed the data, interpreted the results and wrote the first draft of the manuscript. OJT supervised the

study, contributed to the study design, data acquisition**,** and interpretation of the results, and critically revised the manuscript. OSA contributed to data acquisition and spatial analysis and reviewed the manuscript. AOA contributed to the interpretation of the results and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

REFERENCES

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63. *https://doi.org/10.3322/caac.21834*.
2. Fu M, Peng Z, Wu M, Lv D, Lyu S, Li Y. Assessing the African burden of breast cancer: a demographic analysis using Global Cancer Observatory 2022. Eur J Surg Oncol. 2025;51(5):109627. *https://doi.org/10.1016/j.ejso.2025.109627*
3. International Agency for Research on Cancer. Nigeria fact sheet: Global Cancer Observatory (GLOBOCAN 2022) [Internet]. World Health Organization; 2024. Available from: *https://gco.iarc.who.int/media/globocan/factsheets/populations/566-nigeria-fact-sheet.pdf*
4. Anyigba CA, Awandare GA, Paemka L. Breast cancer in sub-Saharan Africa: the current state and uncertain future. Exp Biol Med (Maywood). 2021;246(12):1377-87. *https://doi.org/10.1177/15353702211006047*.
5. Ogunniyi TJ, Fatokun BS, Isah KO, Abdulbaki A, Emiola AR, Batisani K. Current status of cancer diagnosis and treatment in Nigeria. Health Sci Rep. 2025;8:e70877. https://doi.org/10.1002/hsr2.70877.
6. Hoxha I, Sadiku F, Hoxha L, Nasim M, Buteau MAC, Grezda K, et al. Breast cancer and lifestyle factors: umbrella review. Hematol Oncol Clin North Am. 2024;38(1):137-70. *https://doi.org/10.1016/j.hoc.2023.07.005*.
7. Løyland B, Sandbekken IH, Grov EK, Utne I. Causes and risk factors of breast cancer, what do we know for sure? An evidence synthesis of systematic reviews and meta-analyses. Cancers (Basel). 2024;16(8):1583. *https://doi.org/10.3390/cancers16081583*.
8. Tzenios N, Tazanios M, Chahine M. The impact of BMI on breast cancer—an updated systematic review and meta-analysis. Medicine (Baltimore). 2024;103(5):e36831. *https://doi.org/10.1097/MD.0000000000036831*.
9. Chen H, Yuan M, Quan X, Chen D, Yang J, Zhang C, et al. The relationship between central obesity and risk of breast cancer: a dose-response meta-analysis of 7,989,315 women. Front Nutr. 2023;10:1236393. *https://doi.org/10.3389/fnut.2023.1236393*.

10. Sohi I, Rehm J, Saab M, Virmani L, Franklin A, Sánchez G, et al. Alcoholic beverage consumption and female breast cancer risk: a systematic review and meta-analysis of prospective cohort studies. Alcohol Clin Exp Res (Hoboken). 2024;48(12):2222-41. *https://doi.org/10.1111/acer.15493*.

11. Torres-de la Roche LA, Acevedo-Mesa A, Lizarazo IL, Devassy R, Becker S, Krentel H, et al. Hormonal contraception and the risk of breast cancer in women of reproductive age: a meta-analysis. Cancers (Basel). 2023;15(23):5624. *https://doi.org/10.3390/cancers15235624*.

12. Sayon-Orea C, Romanos-Nanclares A, Bullón-Vela V, Chiva-San Roman I, Caiado B, Ondarra-Erdozain M, et al. Effect of duration of hormonal contraceptive use on breast cancer risk: a systematic review and meta-analysis of cohort studies. Maturitas. 2025;202:108750. *https://doi.org/10.1016/j.maturitas.2025.108750*.

13. National Bureau of Statistics (NBS) [Nigeria]. Nigeria Multidimensional Poverty Index (2022) [Internet]. Abuja: National Bureau of Statistics; 2022. Available from: *https://www.nigerianstat.gov.ng/pdfuploads/NIGERIA%20MULTIDIMENSIONAL%20POVERTY%20INDEX%20SURVEY%20RESULTS%202022.pdf*

14. Wang R, Wang P, Zhou Y, Wang Y, Xu C, Wang Z, et al. Association between long-term ambient air pollution exposure and the incidence of breast cancer: a meta-analysis based on updated evidence. Ecotoxicol Environ Saf. 2025;289:117472. *https://doi.org/10.1016/j.ecoenv.2024.117472*.

15. Hvidtfeldt UA, Poulsen AH, Sørensen M, Pedersen JE, Ketzel M, Brandt J, et al. Air pollution with NO_2, $PM_{2.5}$ and elemental carbon in relation to risk of breast cancer: a nationwide case-control study from Denmark. Environ Res. 2023;216:114740. *https://doi.org/10.1016/j.envres.2022.114740*.

16. Heck JE, He D, Wing SE, Ritz B, Carey CD, Yang J, et al. Exposure to outdoor ambient air toxics and risk of breast cancer: the Multiethnic Cohort. Int J Hyg Environ Health. 2024;259:114362. *https://doi.org/10.1016/j.ijheh.2024.114362*.

17. Lai KY, Sarkar C, Ni MY, Cheung LWT, Gallacher J, Webster C. Exposure to light at night (LAN) and risk of breast cancer: a systematic review and meta-analysis. Sci Total Environ. 2021;764:143159. *https://doi.org/10.1016/j.scitotenv.2020.143159*.

18. Luo Z, Liu Z, Chen H, Liu Y, Tang N, Li H. Light at night exposure and risk of breast cancer: a meta-analysis of observational studies. Front Public Health. 2023;11:1276290. *https://doi.org/10.3389/fpubh.2023.1276290*.

19. Agay-Shay K, Michael Y, Lensky IM, Levy I, Blank M, Kayyal-Tarabeia I. Residential greenness and lower breast and prostate cancer incidence: evidence from a retrospective cohort study of 977,644 participants from Israel. Sci Total Environ. 2024;918:170631. *https://doi.org/10.1016/j.scitotenv.2024.170631*.

20. Bhattacharya AA, Allen E, Umar N, Audu A, Felix H, Schellenberg J, et al. Improving the quality of routine maternal and newborn data captured in primary health facilities in

Gombe State, Northeastern Nigeria: a before-and-after study. BMJ Open. 2020;10:e038174. *https://doi.org/10.1136/bmjopen-2020-038174*.

21. Oladoyin V, Adedini S, Ijadunola K, Ogunwemimo H, Folorunso O, Chukwu E, et al. The assessment of routine health information system performance towards improvement of quality of reproductive, maternal, newborn, child and adolescent health services in Ondo and Ekiti States, Nigeria. PLoS One. 2025;20(1):e0318010. *https://doi.org/10.1371/journal.pone.0318010*.
22. Jedy-Agba E, Curado MP, Ogunbiyi O, Oga E, Fabowale T, Igbinoba F, et al. Cancer incidence in Nigeria: a report from population-based cancer registries. Cancer Epidemiol. 2012;36(5):e271-8. *https://doi.org/10.1016/j.canep.2012.04.007*.
23. Zubairu IH, Balogun MS. Population-based cancer registries in Nigeria and the National Cancer Control Programme. Ecancermedicalscience. 2023;17:1592. *https://doi.org/10.3332/ecancer.2023.1592*.
24. Soleimani M, Ayyoubzadeh SM, Jalilvand A, Ghazisaeedi M. Exploring the geospatial epidemiology of breast cancer in Iran: identifying significant risk factors and spatial patterns for evidence-based prevention strategies. BMC Cancer. 2023;23:1219. *https://doi.org/10.1186/s12885-023-11555-1*.
25. OECD, UCLG. Nigeria: country profile [Internet]. World Observatory on Subnational Government Finance and Investment (SNG-WOFI); [cited 2026 Oct 25]. Available from: *https://www.sng-wofi.org/country_profiles/nigeria.html*
26. National Population Commission (NPC) [Nigeria], ICF. Nigeria Demographic and Health Survey 2024. Abuja, Nigeria and Rockville, Maryland, USA: NPC and ICF; 2025.
27. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2024 [Internet]. New York: United Nations; 2024 [cited 2026 Oct 25]. Available from: *https://population.un.org/wpp/*
28. Nigerian Information Portal. About Nigeria [Internet]. [cited 2026 Oct 25]. Available from: *https://nigeria.gov.ng/?p=45*
29. Africa Health Observatory Platform, WHO Regional Office for Africa. Organization and governance of the health system. In: Nigeria Health Systems and Services Profile [Internet]. 2025 [cited 2026 Oct 25]. Available from: *https://ahop.aho.afro.who.int/download/organization-and-governance-of-the-health-system-nigeria-health-system-and-services-profile*
30. Jedy-Agba EE, Oga EA, Odutola M, Abdullahi Y, Popoola A, Achara P, et al. Developing national cancer registration in developing countries: case study of the Nigerian National System of Cancer Registries. Front Public Health. 2015;3:186. *https://doi.org/10.3389/fpubh.2015.00186*.
31. National Institute for Cancer Research and Treatment. Cancer control capacity and needs assessment in Nigeria: imPACT review report [Internet]. 2023. Available from:

https://www.nicrat.gov.ng/wp-content/uploads/2023/08/Nigeria-imPACT-Review-Report.pdf

32. Global Data Lab. Subnational Human Development Index (SHDI) [dataset] [Internet]. Global Data Lab; 2022. Available from: *https://globaldatalab.org/shdi/*

33. Van Geffen J, Eskes H, Compernolle S, Pinardi G, Verhoelst T, Lambert JC, et al. Sentinel-5P TROPOMI NO_2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data. Atmos Meas Tech. 2022;15(7):2037-60. *https://doi.org/10.5194/amt-15-2037-2022*.

34. Copernicus Atmosphere Monitoring Service. CAMS global atmospheric composition forecasts and reanalysis [Internet]. 2026. Available from: *https://atmosphere.copernicus.eu/*

35. Wan Z, Hook S, Hulley G. MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061 [dataset]. NASA EOSDIS Land Processes DAAC; 2021. *https://doi.org/10.5067/MODIS/MOD11A1.061*.

36. Bondarenko M, Kerr D, Sorichetta A, Tatem AJ. Census/projection-disaggregated gridded population datasets for 189 countries in 2020 using built-settlement growth model (BSGM) outputs [dataset]. WorldPop, University of Southampton; 2020. *https://doi.org/10.5258/SOTON/WP00684*.

37. Food and Agriculture Organization of the United Nations. The Global Administrative Unit Layers (GAUL) 2024: technical guidelines. FAO; 2025. *https://doi.org/10.4060/cd4262en*.

38. Chen Y. Spatial autocorrelation equation based on Moran's index. Sci Rep. 2023;13:19296. *https://doi.org/10.1038/s41598-023-45947-x*.

39. Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest Geogr. 2011;30(2):87-93. *https://doi.org/10.2478/v10117-011-0021-1*.

40. Ojeleke O, Groot W, Bonuedi I, Pavlova M. The impact of armed conflicts on health-care utilization in Northern Nigeria: a difference-in-differences analysis. World Med Health Policy. 2022;14:624-64. *https://doi.org/10.1002/wmh3.501*.

41. Vanderpuye V, Hammad N, Ndlovu N, Simonds H, Goliath R, Othieno-Abinya NA, et al. Real-world challenges for patients with breast cancer in sub-Saharan Africa: a retrospective observational study of access to care in Ghana, Kenya and Nigeria. BMJ Open. 2021;11(3):e041900. *https://doi.org/10.1136/bmjopen-2020-041900*.

42. Mathelin C, Ame S, Anyanwu SNC, Avisar E, Boubnider WM, Breitling K, et al. Breast cancer management during the COVID-19 pandemic: the Senologic International Society survey. Eur J Breast Health. 2021;17(2):188-96. *https://doi.org/10.4274/ejbh.galenos.2021.2021-1-4*.

43. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36

cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. *https://doi.org/10.3322/caac.21492*.

44. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol. 2012;36(3):237-48. *https://doi.org/10.1016/j.canep.2012.02.007*.

45. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-49. *https://doi.org/10.3322/caac.21660*.

46. Ordinioha B, Brisibe S. The human health implications of crude oil spills in the Niger Delta, Nigeria: an interpretation of published studies. Niger Med J. 2013;54(1):10-6. *https://doi.org/10.4103/0300-1652.108887*.

47. Makinde OA, Williams A, Adeleke O, Ohadi EM, Dieng AA, Osika JS. Assessment of the routine health management information system in Oyo State, Federal Republic of Nigeria. Health Systems 20/20 Project, Abt Associates Inc.; 2012.

48. AbouZahr C, Boerma T. Health information systems: the foundations of public health. Bull World Health Organ. 2005;83(8):578-83. Available from: *https://scielosp.org/article/bwho/2005.v83n8/578-583*

49. Mutale W, Chintu N, Amoroso C, Awoonor-Williams JK, Phillips JF, Baynes C, et al. Improving health information systems for decision making across five sub-Saharan African countries: implementation strategies from the African Health Initiative. BMC Health Serv Res. 2013;13(Suppl 2):S9. *https://doi.org/10.1186/1472-6963-13-S2-S9*.

50. Farnham A, Loss G, Lyatuu I, Cossa H, Kulinkina AV, Winkler MS. A roadmap for using DHIS2 data to track progress in key health indicators in the Global South: experience from sub-Saharan Africa. BMC Public Health. 2023;23:1030. *https://doi.org/10.1186/s12889-023-15979-z*.

51. Roheel A, Khan A, Anwar F, Akbar Z, Akhtar MF, Khan MI, et al. Global epidemiology of breast cancer based on risk factors: a systematic review. Front Oncol. 2023;13:1240098. *https://doi.org/10.3389/fonc.2023.1240098*.

52. GBD 2023 Breast Cancer Collaborators. Global, regional, and national burden of breast cancer among females, 1990-2023, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2023. Lancet Oncol. 2026;27(3):302-26. *https://doi.org/10.1016/S1470-2045(25)00730-2*.

53. Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995;27(2):93-115. *https://doi.org/10.1111/j.1538-4632.1995.tb00338.x*.

54. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geogr Anal. 1992;24(3):189-206. *https://doi.org/10.1111/j.1538-4632.1992.tb00261.x*.

55. Liu T, Chen R, Zheng R, Li L, Wang S. Household air pollution from solid cooking fuel combustion and female breast cancer. Front Public Health. 2021;9:677851. *https://doi.org/10.3389/fpubh.2021.677851*.

56. Ibrahim AO, Omonijo A, Agbesanwa TA, Alabi AK, Elegbede OT, Olusuyi KM, et al. A 14-year analysis of breast cancer risk factors and its determinants of mortality in rural southwestern Nigeria. Clin Med Insights Oncol. 2024;18. *https://doi.org/10.1177/11795549241288197*.

57. Omisore AD, Sutton EJ, Akinola RA, Towoju AG, Akhigbe A, Ebubedike UR, et al. Population-level access to breast cancer early detection and diagnosis in Nigeria. JCO Glob Oncol. 2023;9:e2300093. *https://doi.org/10.1200/GO.23.00093*.

58. Mane N, Fouqani A, Mrah S, Omari M, Bouaddi O, Faure E, et al. Obesity and risk of pre- and postmenopausal breast cancer in Africa: a systematic review. Curr Oncol. 2025;32(3):167. *https://doi.org/10.3390/curroncol32030167*.

59. Prajapati N, Cordina-Duverger E, Boileau A, Faure E, Guénel P. Exposure to outdoor artificial light at night and breast cancer risk: a population-based case-control study in two French departments (the CECILE study). Front Environ Health. 2023;2:1268828. *https://doi.org/10.3389/fenvh.2023.1268828*.

60. Sweeney MR, Nichols HB, Jones RR, Olshan AF, Keil AP, Engel LS, et al. Light at night and the risk of breast cancer: findings from the Sister Study. Environ Int. 2022;169:107495. *https://doi.org/10.1016/j.envint.2022.107495*.

61. Fayehun O, Apenteng P, Usman U, Adebayo K, Owoaje E, Sartori J, et al. Diagnosis of cancer in the South and North of Nigeria: duration and causes of delay. BMC Health Serv Res. 2025;25:738. *https://doi.org/10.1186/s12913-025-12707-8*.

62. U.S. Environmental Protection Agency. Formaldehyde (CASRN 50-00-0): toxicological review (IRIS assessment) [Internet]. 2024. Available from: *https://iris.epa.gov/*

63. Agodirin O, Chijioke C, Mustapha F, Rahman G, Olatoke S, Olaogun J, et al. Nigeria's 6-year (2018–2023) stage distribution of breast cancer at diagnosis: a systematic review and meta-analysis. Ecancermedicalscience. 2025;19:1899. *https://doi.org/10.3332/ecancer.2025.1899*.

64. Li J, Xie Y, Xu J, Zhang C, Wang H, Huang D, et al. Association between greenspace and cancer: evidence from a systematic review and meta-analysis of multiple large cohort studies. Environ Sci Pollut Res Int. 2023;30(39):91140-57. *https://doi.org/10.1007/s11356-023-28461-5*.

65. Nakkazi E. Not just a climate issue: cutting cancer rates through cleaner cooking fuels in Africa [Internet]. Cancerworld; 2024 Feb 9. Available from: *https://cancerworld.net/not-just-a-climate-issue-cutting-cancer-cooking-fuels-in-africa/*

66. James JO, Adewumi T, Durodola OM, Okonkwo RN, Abimbola OJ. Multi-sensor satellite data assessment of spatio-temporal dynamics of air pollution in Abuja FCT and

adjoining states in Nigeria. Lafia J Sci Ind Res. 2025;3(2):122-36. *https://doi.org/10.62050/ljsir2025.v3n2.632*.

67. International Agency for Research on Cancer. Alcohol use and breast cancer: a critical review [Internet]. Lyon: IARC; 2021. Available from: *https://www.iarc.who.int/reference/alcohol-use-and-breast-cancer-a-critical-review*

68. Niemeyer Hultstrand J, Gemzell-Danielsson K, Kallner HK, Lindman H, Wikman P, Sundström-Poromaa I. Hormonal contraception and risk of breast cancer and breast cancer in situ among Swedish women 15–34 years of age: a nationwide register-based study. Lancet Reg Health Eur. 2022;21:100470. *https://doi.org/10.1016/j.lanepe.2022.100470*.

69. Rybnikova N, Haim A, Portnov BA. Artificial light at night (ALAN) and breast cancer incidence worldwide: a revisit of earlier findings with analysis of current trends. Chronobiol Int. 2015;32(6):757-73. *https://doi.org/10.3109/07420528.2015.1043369*.

70. James P, Bertrand KA, Hart JE, Schernhammer ES, Tamimi RM, Laden F. Outdoor light at night and breast cancer incidence in the Nurses' Health Study II. Environ Health Perspect. 2017;125(8). *https://doi.org/10.1289/EHP935*.

Supplementary Files

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

- SUPPLEMENTARYINFORMATIONFILE.docx