

Spatial Distribution and Clinicopathological Factors Associated with Late-Stage Breast Cancer in Kelantan, Malaysia: A Geographic Information Systems (GIS)-Based Analysis

Aishah Md Salleh

Department of Surgery, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia.

Wan Zainira Wan Zain

Department of Surgery, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia. Department of Surgery, Hospital Pakar Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia.

Zalina Zahari

Faculty of Pharmacy, Universiti Sultan Zainal Abidin (UniSZA), Besut Campus, Besut, Terengganu, Malaysia.

Maya Mazuwin Yahya

Department of Surgery, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia. Department of Surgery, Hospital Pakar Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia.

Ahmad Filza Ismail

Department of Community Medicine, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia.

Introduction: Breast cancer remains the most common cancer in Malaysia, with 47.9% of cases diagnosed at late stages (stage III and IV). Geographic Information Systems (GIS) are used to analyse spatial data and understand the distribution of diseases. This study aimed to assess the spatial distribution of breast cancer and identify clinicopathological factors associated with metastatic disease (stage IV) among late-stage patients.

Materials and Methods: This retrospective study included female patients with histopathologically confirmed breast cancer. A total of 224 patients from Hospital Pakar Universiti Sains Malaysia (HPUSM), Kelantan, Malaysia were analysed. This study used GIS to assess spatial distribution patterns of breast cancer cases and logistic regression to determine clinicopathological factors associated with metastatic versus locally advanced (stage III) disease.

Results: Spatial analysis revealed significant clustering of late-stage breast cancer cases (NNR: 0.44, Z-score: -13.18, $P < 0.001$). Late-stage patients were more likely to reside within 10 km of the nearest available hospitals (75.2%) compared to early-stage patients (45.1%) ($\chi^2(1, N = 224) = 19.47, P < 0.001$). Progesterone receptor (PR)-positive status was associated with 90% lower odds to present with stage IV disease rather than stage III compared to PR-negative patients (Adjusted odds ratio (AOR): 0.10, 95% CI: 0.01-0.89, $P = 0.039$).

Conclusions: Results from our study highlighted areas with a higher concentration of advanced disease and suggested that patients living closer to healthcare facilities may present with more advanced disease. Furthermore, PR positivity was identified as a significant predictor of metastatic disease among late-stage patients. These findings underscore the potential of GIS to guide hotspot-targeted screening initiatives, such as mobile mammography in high-risk areas, and highlight the value of PR status as a marker for risk stratification in late-stage patients, informing both clinical decision-making and targeted

public health interventions.

Introduction

Breast cancer remains the most common malignancy among women worldwide and one of the leading causes of cancer-related mortality. According to GLOBOCAN 2022, breast cancer accounted for 12.5% of all newly diagnosed cancers and 6.8% of global cancer deaths [1]. In Malaysia, it continues to be the predominant cancer among women, representing 34.1% of all female cancers as reported in the Malaysian National Cancer Registry Report (MNCRR 2012–2016) [2]. The age-standardised incidence rate increased from 31.1 per 100,000 population in 2007–2011 to 34.1 in 2012–2016. Among cases with known staging information, 47.9% were diagnosed at late stages (III and IV), an increase from 43.2% in the previous report [2]. Despite continuous efforts in awareness campaigns and screening programmes, a considerable proportion of Malaysian women are still diagnosed at advanced stages of the disease. This pattern of delayed presentation is commonly observed in middle-income countries and contributes to poorer survival outcomes compared with high-income nations [3-5].

Geographic Information Systems (GIS) are powerful tools for gathering, storing, manipulating, and visualising geographically referenced data. In epidemiological research, GIS enables the measurement of spatial patterns and relationships between health-related variables, offering insights that complement traditional statistical analyses [6-10]. The application of GIS in health has been widely recognized for its utility in understanding geographical determinants of health outcomes, particularly in oncology [11-14]. In cancer research, GIS has been increasingly applied to visualise disease distribution, identify spatial clustering, and evaluate accessibility to healthcare facilities [9, 10, 15-17]. These capabilities support a deeper understanding of how geographic and sociodemographic factors influence disease stage at diagnosis and treatment outcomes, ultimately guiding data-driven public health strategies and equitable resource allocation.

Breast cancer is a heterogeneous and biologically complex disease, encompassing multiple subtypes with distinct molecular and clinical characteristics. Variations in treatment response and prognosis are influenced by tumour size, lymph node involvement, histological grade, patient age, and key biomarkers such as oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) status [18-20]. While traditional histopathological classification remains essential, it may not fully capture the underlying molecular alterations driving tumour behaviours [5, 18]. Tumours with similar histological features can exhibit markedly different clinical behaviours due to these biological differences, underscoring the need to understand all factors, including novel therapeutic agents and adjuvant treatments, that influence clinical outcomes [19-22].

From a clinical standpoint, hormone receptor profiling provides important insight into tumour behaviour and disease trajectory. In particular, PR positivity is often associated with less aggressive phenotypes and better response to endocrine therapy [19, 20]. Understanding how these biological markers interact with geographic and healthcare accessibility factors may offer a more integrated perspective on the determinants of late-stage breast cancer [8, 9, 16].

Despite significant advances in screening and treatment, gaps remain in understanding the geographic distribution and clinicopathological determinants of late-stage breast cancer presentation in Malaysia. Applying GIS allows for the visualization of spatial patterns and identification of potential clusters of advanced disease, which may inform more effective early detection and outreach strategies.

Therefore, this study aimed to assess the spatial distribution of breast cancer cases using GIS and to identify clinicopathological factors associated with metastatic disease (stage IV) compared to

locally advanced disease (stage III) among the late-stage breast cancer patients treated at Hospital Pakar Universiti Sains Malaysia (HPUSM). Insights from this study may support the development of targeted screening initiatives and optimized resource allocation for high-risk populations.

Materials and Methods

Study Design and Setting

This was a retrospective cross-sectional study that included patients with breast cancer who sought treatment at the Department of Surgery, HPUSM, Kelantan, Malaysia from January 1, 2018 to December 31, 2020. HPUSM is an academic, tertiary care center, and one of the largest hospitals in Kelantan.

Participants

During the study period, a total of 224 female patients with histopathologically confirmed breast cancer were sampled from the HPUSM record office and met the eligibility criteria. The inclusion criteria for this study were female patients who sought treatment at HPUSM during the study period and had histopathologically confirmed breast cancer. Patients with incomplete medical records and male breast cancer patients were excluded. An ethical approval was obtained from the Human Research Ethics Committee USM (HREC), Kelantan, Malaysia (Reference No: USM/JEPeM/21060497).

Data Collection

The records of patients diagnosed with breast cancer at HPUSM between 2018 and 2020 were identified from the HPUSM record office, and their details (address, gender, age, ethnicity, and clinical data including cancer stage, affected breast side, oestrogen receptor, progesterone receptor, biopsy, and histopathology) were collected from medical records. All the relevant information was recorded in the data collection form to reduce bias. All the collected data were kept confidential, accessible only to the researcher and co-researchers. The data were stored in Microsoft Excel format with password protection, while paper copies of the data and reports were securely archived in the Department of Surgery, HPUSM. The coordinates were obtained via Google Maps using the patients' addresses.

Statistical Analysis

GIS software, including ArcGIS version 10.7 and QGIS version 3.12.1, was used for spatial and zonal analysis, including Z-score calculations [9]. Spatial analysis methods applied in this study included the Average Nearest Neighbour (ANN) and Hot Spot Analysis (Getis-Ord G_i^*). The Average Nearest Neighbour determined whether case distributions were clustered or random, while Getis-Ord G_i^* provided a Z-score to indicate spatial clustering of high or low values. A higher Z-score represented stronger hot spot clustering. Both analyses were performed to achieve the study objectives. The geographical coordinates were projected using the GDM2000 coordinate system, and Euclidean distance (straight-line distance) was used as the distance metric for all spatial calculations.

Mean and standard deviation (SD) were computed for numerical variables (age), and for categorical variables (affected breast side, oestrogen receptor, progesterone receptor, biopsy, histopathology

examination, breast cancer stage and distance from hospital), frequency and percentages were also computed. Logistic regression analysis was performed to identify clinicopathological factors associated with metastatic disease (stage IV) among late-stage patients. The multiple logistic regression analysis was conducted exclusively on the sub-group of late-stage patients (N = 164), comparing metastatic disease (stage IV) as the dependent variable against locally advanced disease (stage III) as the reference category. Independent variables were possible risk factors for metastatic disease among late-stage patients included age, affected breast side, oestrogen receptor, progesterone receptor, biopsy, and histopathology. Initially, simple logistic regression was conducted to obtain the crude odds ratio of the predictors, followed by multiple logistic regression to determine the adjusted odds ratio. The model was evaluated for goodness-of-fit. Data analyses were performed using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA). The limit of significance was set at 0.05.

Sample Size Estimation

The sample size was estimated based on primary objective using sample size calculator, [23] at https://wnarifin.github.io/ssc_web.html based on two proportion formula to obtain the appropriate sample size for the clinicopathological factors associated with late stages breast cancer. The calculation was based on a study power of 80%, a significance level (α) of 0.05, and an expected dropout rate of 10%. The proportion values (P0 and P1) were determined based on the Malaysian National Cancer Registry Report 2012–2016 [2]. The baseline proportion (P0) was set according to the late-stage prevalence reported nationally (47.9%). The value for P1 was hypothesized to detect a 12% absolute difference in the prevalence of a specific risk factor between the early- and late-stage groups. Based on these parameters, the minimum required sample size was calculated to be 188. Our study, however, adopted a consecutive sampling strategy over the study period (January 1, 2018, to December 31, 2020), which resulted in the inclusion of all eligible patients (N = 224) who sought treatment at HPUSM during that time frame. The final sample size was thus dictated by the availability of patient records during the study period, exceeding the minimum requirement of 188, which provides adequate statistical power for the study objective.

Results

Patient Demographics and Clinical Characteristics

A total of 224 breast cancer patients who received treatment at the Department of Surgery, HPUSM from 2018 to 2020 had complete medical records. All patients were included in the study, meeting the required sample size of 188. The mean age of breast cancer patients was 52.2 years (SD = 12.01). Breast cancer occurred predominantly in the left breast (49.1%), followed by the right (46.4%) and bilateral cases (4.5%). More than half of the patients (66.5%, n = 149) tested positive for oestrogen receptor, while 54.0% (n = 121) were positive for progesterone receptor. Histopathological examination revealed that the most common biopsy finding was invasive carcinoma of no special type (NST) (69.6%), followed by other subtypes such as invasive ductal carcinoma (4.0%) and invasive lobular carcinoma (2.2%). As shown in Table 1, a majority (58.9%) of patients were diagnosed at stage IV, while 17.4% were at stage II, 14.3% at stage III, and only 9.4% at stage I.

Variable	All patients (N = 224)
Age	52.2 (12.01)*
Affected breast side	
Left	110 (49.1)
Right	104 (46.4)
Bilateral	10 (4.5)

Oestrogen receptor	
Negative	75 (33.4)
Positive	149 (66.5)
Progesterone receptor	
Negative	103 (45.9)
Positive	121 (54.0)
Biopsy	
Invasive carcinoma of no special type (NST)	180 (80.4)
Invasive lobular carcinoma	6 (2.7)
Invasive ductal carcinoma	12 (5.4)
Mucinous carcinoma	6 (2.7)
Others	20 (8.9)
Histopathology examination	
Invasive carcinoma of no special type (NST)	156 (69.6)
Invasive lobular carcinoma	5 (2.2)
Invasive ductal carcinoma	9 (4.0)
Mucinous carcinoma	5 (2.2)
Others	49 (21.8)
Breast cancer stage	
Stage I	21 (9.4)
Stage II	39 (17.4)
Stage III	32 (14.3)
Stage IV	132 (58.9)

Table 1. Demographics and Clinical Characteristics of Breast Cancer Patients (N=224).

Values are presented as frequency (percentage). *Mean and standard deviation (SD)

Distribution of Breast Cancer Cases

The Hot Spot Analysis (Figure 1) maps the spatial distribution of breast cancer cases, revealing a high concentration in the northeast and northwest regions of Kelantan with 90% confidence.

Figure 1. Hot Spot Analysis Showing Spatial Distribution of All Breast Cancer Cases at HPUSM, with Clustering in Northern Kelantan.

This spatial pattern aligns to the spatial distribution of late-stage breast cancer cases, as shown in Figure 2 a), where localized hot spots indicate higher disease concentration. In contrast, due to the sparse distribution of early-stage breast cancer cases, a distinct hot spot could not be identified, as illustrated in Figure 2 b).

Figure 2. a) Hot Spot Analysis of Late-Stage Breast Cancer Patients in HPUSM, b) Hot Spot Analysis of Early-Stage Breast Cancer Patients in HPUSM.

Figures 3 a) and b), generated using ArcGIS, illustrate the number of patients residing within a 10 km radius of the nearest available hospitals.

Figure 3. a) Distribution of Late-Stage Breast Cancer Patients in HPUSM within a 10 km Radius of Hospitals, b)

Distribution of Early-Stage Breast Cancer Patients in HPUSM within a 10 km Radius of Hospitals.

As shown in Table 2, there was a significant association between breast cancer stages and proximity (within 10 km) to hospitals, $\chi^2 (1, N = 224) = 19.47, P < 0.001$.

Variable	Breast cancer stage		x2 (df)	P-value†
	Early-stage	Late-stage		
Distance from hospital				
Within 10 km	32 (45.1%)	115 (75.2%)	19.47 (1)	< 0.001
More than 10 km	39 (54.9%)	38 (24.8%)		

Table 2. Association between Breast Cancer Stages and Proximity to the Nearest Hospital (10 km Radius).

†Pearson's chi-square test

Late-stage breast cancer patients were more likely to reside within 10 km of a hospital (75.2%) compared to early-stage patients (45.1%).

The spatial distribution of breast cancer cases in this study was categorised into early-stage and late-stage breast cancer. Interestingly, early-stage breast cancer cases exhibited a random pattern (NNR: 0.92, Z-score: -1.05, P = 0.292), as shown in Figure 4 b), compared to the distribution in Figure 4 a) (NNR: 0.44, Z-score: - 13.18, P < 0.001).

Figure 4. a) Average Nearest Neighbour Summary for Late-Stage Breast Cancer Patients in HPUSM, b) Average Nearest Neighbour Summary for Early-Stage Breast Cancer Patients in HPUSM.

Factors Associated with Metastatic (Stage IV) Compared to Locally Advanced (Stage III) Disease among Late-Stage Breast Cancer Patients

Table 3 shows univariable and multivariable analyses on the predictors of metastatic disease among late-stage patients.

Variable	Simple logistic regression	P-value	Multiple logistic regression*	P-value
	Crude OR (95% CI)		Adjusted OR (95% CI)	
Age	0.99 (0.96, 1.03)	0.753	0.99 (0.96, 1.03)	0.711
Affected breast side				
Left	1.33 (0.61, 2.91)	0.477	1.33 (0.56, 3.18)	0.530
Right (Ref)	1		1	
Oestrogen receptor				
Negative (Ref)	1		1	
Positive	1.06 (0.46, 2.47)	0.888	7.62 (0.87, 66.73)	0.067
Progesterone receptor				
Negative (Ref)	1		1	
Positive	0.58 (0.26, 1.30)	0.185	0.10 (0.01, 0.89)	0.039
Biopsy				
Invasive carcinoma of no special type (NST)	1.37 (0.35, 5.41)	0.653	3.00 (0.18, 49.71)	0.444
Invasive lobular carcinoma	—	—	—	—

Invasive ductal carcinoma	1.67 (0.14, 20.58)	0.69	4.20 (0.13, 138.69)	0.422
Mucinous carcinoma	1.00 (0.07, 13.64)	1	—	—
Others (Ref)	1		1	
Histopathology examination				
Invasive carcinoma of no special type (NST)	0.92 (0.24, 3.54)	0.899	0.61 (0.05, 7.20)	0.696
Invasive lobular carcinoma	—	—	—	—
Invasive ductal carcinoma	1.64 (0.14, 19.39)	0.696	1.79 (0.10, 31.02)	0.689
Mucinous carcinoma	0.27 (0.03, 2.83)	0.276	—	—
Others (Ref)	1		1	

Table 3. Predictors of Metastatic Disease (Stage IV) compared to Locally Advanced Disease (Stage III) in the Late-Stage Subgroup from Simple and Multiple Logistic Regression (N = 164).

OR, odds ratio; CI, confidence interval; Ref, reference category; *All variables were included in the multiple logistic regression. The model reasonably fits well. Model assumptions are met. There are no interaction and multicollinearity problems. Model fitness was confirmed by the non-significant Hosmer-Lemeshow test ($P > 0.05$).

Using simple logistic regression, none of the variables showed a statistically significant association with metastatic disease among late-stage patients. In this dataset, all variables were included in the multiple logistic regression. Multiple logistic regression showed that progesterone receptor status was the predictor of metastatic disease among late-stage patients. The multiple logistic regression analysis was conducted exclusively on the late-stage patient subgroup ($N = 164$), comparing metastatic disease (stage IV) as the dependent variable against locally advanced disease (stage III) as the reference category. The overall model demonstrated adequate fit, confirmed by a non-significant Hosmer-Lemeshow test ($P > 0.05$). After adjustment for age, affected breast side, oestrogen receptor, biopsy, and histopathology, the result showed that PR-positive patients had 90% lower odds of presenting with metastatic (stage IV) disease compared to PR-negative patients (Adjusted odds ratio (AOR): 0.10, 95% CI: 0.01-0.89, $P = 0.039$). Age ($P = 0.711$), breast side ($P = 0.530$), and oestrogen receptor status ($P = 0.067$) were not significantly associated with metastatic disease among late-stage patients.

Discussion

This study utilized a GIS-based analysis to assess the spatial distribution of breast cancer in Kelantan and identify factors associated with metastatic disease (stage IV) compared to locally advanced disease (Stage III) among patients at HPUSM. This study revealed that most breast cancer cases were clustered in the northern part of Kelantan, where HPUSM is located, consistent with a previous spatial breast cancer study in Kelantan [2]. Furthermore, Hot Spot Analysis found that late-stage breast cancer cases were predominantly clustered in the northern region of Kelantan. A similar study conducted in Penang demonstrated comparable findings, suggesting that major city locations contribute to the concentration of cases [9].

The mean age at diagnosis in this study was 52.2 years, indicating a relatively younger mean age compared to census data reported in similar studies [9, 17, 24]. This finding is consistent with the known pattern of breast cancer in Asia, where a higher proportion of cases are diagnosed in pre-menopausal women compared to Western nations. This earlier onset is likely attributable to a combination of factors, including potential genetic predisposition, differences in parity and

reproductive patterns, and recognised epidemiological trends among Malay women in Malaysia and Southeast Asia [5, 25]. The clinical implication of this earlier age of onset is critical, as these younger patients often present with more aggressive disease subtypes and face longer periods of potential life-years lost.

Previous study has concluded that the overall 5-year survival rate of breast cancer patients in Malaysia remains lower compared to developed nations [26]. This disparity in survival, coupled with the tendency for diagnosis at a younger mean age and later stage (as demonstrated by our Stage IV vs. Stage III analysis), underscores an urgent need for targeted public health strategies and enhanced genetic screening programs aimed at high-risk populations in Malaysia. The spatial clusters identified by our GIS analysis should be the immediate focus of these efforts.

One of the important determinants of carcinoma survival is early detection, which influenced by disease awareness and the uptake of screening [27]. Notably, even among high-risk women, a cross-sectional study reported that the majority had poor knowledge about breast cancer risk factors. Women with a family history of breast cancer probably did not recognise their increased risk and, consequently, presented with the same disease stage as those with no family history [3].

This study provided valuable insights into the comparative distribution of early- and late-stage breast cancer patients in a single center. Late-stage presentation has been attributed to strong beliefs in traditional medicine, negative disease perception, poverty, poor education, fear, and denial. A similar pattern of results was obtained in Sabah, patients presenting with advanced disease were also of lower socioeconomic status, limited formal education, and predominantly rural residence [7]. Breast cancer survival depends on prognostic factors, such as stage at first diagnosis, tumour size, menstrual status, and histopathology [4]. Additionally, survival is influenced by complex underlying factors including population structure (age and ethnicity), socio-economic status, and the availability of an effective healthcare system, including screening programs that enhance early detection of cases and accessibility to high-quality treatment [5].

Building on these factors influencing breast cancer survival, the TNM staging system is widely used to facilitate patient management and enable comparison between countries. Differences in screening practices, poor health-seeking behaviour, and poor treatment compliance may contribute to poorer prognosis among Malays compared to other ethnic groups in Malaysia [8]. Delayed presentation remains very common, as reported by a collaborative study in Malaysia, India, and Hong Kong, where only 5.2% of breast cancer cases were detected through mammographic screening. A study also reported that patients with advanced disease were poor, non-educated, and from rural areas [25]. In contrast, this study mapped more cases of late-stage cases in areas with greater screening accessibility, suggesting that many patients had undergone screening before their diagnosis. The clustered population was located within a 10 km radius of the nearest available hospitals, which could at least provide basic diagnostic methods, including clinical evaluation, radiological imaging, or even histopathological examination [7]. Additionally, the late-stage breast cancer cases were found to be spatially clustered. This geographical clustering carries significant public health implications. The spatial analysis provides the necessary data to inform policymakers where to deploy targeted screening programs, such as mobile mammography units or intensive community awareness campaigns, to intercept the disease at an earlier stage. Furthermore, given the predominantly Malay population in the study area, and literature suggesting poorer prognosis among this ethnic group [8], these identified geographical clusters represent vulnerable areas where targeted intervention is critical to address health disparities and improve overall survival outcomes.

Our results demonstrated that most breast cancer cases were diagnosed at a later stage, with 14.3% at stage III and 58.9% at stage IV. This finding is consistent with other GIS study on colorectal cancer, which further emphasise the need for early detection through mass screening [10] to increase early-stage diagnoses and reduce the incidence of later-stage cases. Comparatively, a study on the spatial distribution of colorectal cancer staging showed variations in stage-specific

incidences [15].

In this study, spatial and temporal analysis using GIS has shown that breast cancer clustering is highest in urban areas, while cases in rural areas follow a random distribution. Unlike conventional statistical methods, GIS-based analysis allows for disease spread mapping, showing whether cases spread towards-urban-agglomeration or towards-rural-regions or both, helping to understand the growth of urban facilities, lifestyle changes, and their adaptation and acceptance in rural areas [28].

Furthermore, spatial analysis may help identify new exposure hypotheses that warrant future epidemiologic investigations with detailed exposure models. The current analyses illustrate the usefulness of spatial and temporal analyses in visualising cancer risk, adjusting for confounding factors, and assessing the statistical significance of location and time. Notably, the results of this study showed that 90% of breast cancer cases are located within a 10 km radius of the nearest hospital, suggesting excellent accessibility to public hospitals. This finding is consistent with a previous study on hospital proximity and disease clustering [6].

The finding that late-stage breast cancer patients were significantly more likely to reside within 10 km of a healthcare facility appears counterintuitive and illustrates a “Proximity Paradox.” In most settings, greater proximity to healthcare facilities is expected to facilitate earlier diagnosis [7-9]. However, in this study, patients living closer to hospitals presented with more advanced disease, which may be explained by several contextual factors. One possible explanation is referral bias, as HPUSM is a tertiary referral centre that receives a high volume of advanced cases from surrounding districts and primary or secondary facilities within its immediate urban catchment area. Similar clustering of late-stage or complex cases near tertiary hospitals has been reported in other Malaysian GIS studies [7, 8]. This referral pattern may therefore inflate the proportion of late-stage cases among populations residing near HPUSM.

Another contributing factor could be delayed health-seeking behaviour among urban residents, despite their physical proximity to healthcare services. Prior studies in Malaysia have documented that women may postpone medical consultation due to denial, fear of diagnosis, competing responsibilities, or lack of social support [25]. Patients living near hospitals may perceive easy access and thus delay presentation until symptoms become severe. Together, these findings suggest that geographic proximity alone does not ensure early detection. Instead, behavioural, psychosocial, and system-level barriers must also be addressed to improve timely diagnosis and reduce late-stage presentation.

Our study investigated the predictors of metastatic disease among late-stage patients using univariable and multivariable analyses. Simple logistic regression did not identify any variables with a statistically significant association with metastatic disease among late-stage patients. However, multiple logistic regression analysis showed that progesterone receptor (PR) status was significantly associated with metastatic disease among late-stage patients. This finding suggests that PR expression may play a protective role in limiting disease advancement, possibly due to its influences on tumour aggressiveness and metastatic potential [18].

Notably, our results contrast with previous studies that have primarily linked age at diagnosis and other clinicopathological variables such as TNM staging, molecular subtypes, and therapeutic options to breast cancer progression [19, 20]. Others have shown that tumours in younger and older patients exhibit biological differences, primarily depended on the presence of oestrogen and progesterone receptor positivity or hormone sensitivity. Tumours that lack hormone sensitivity tend to progress faster and may present at a higher grade at the time of diagnosis [18]. In line with this, in the inferential analysis, we found that PR negativity was independently associated with a higher likelihood of being diagnosed with metastatic disease (stage IV). While statistically significant, the wide confidence interval (95% CI: 0.01- 0.89) indicates the imprecision of this estimate, suggesting caution in interpreting the magnitude of the effect. PR- negative status is classically linked to a

more aggressive, less differentiated tumor phenotype and resistance to endocrine therapy, which aligns with poorer outcomes [29, 30]. Biologically, PR expression acts as a marker of a fully functional Estrogen Receptor (ER) signaling pathway. The presence of PR, which is regulated by ER activity, often correlates with lower tumor proliferation rates and better differentiation (e.g., Luminal A subtype), thereby reducing the likelihood of metastatic spread at diagnosis [31]. Our finding reinforces the clinical utility of hormone receptor profiling, even when assessing differences among late-stage disease, and highlights the need for tailored, aggressive treatment planning for patients presenting with this biological profile.

This study has several limitations that should be considered. First, it was conducted at a single tertiary referral center (HPUSM), may introduce selection bias and as such, the patient population is limited by the local demographic and the expertise of the treating physicians. Given that HPUSM is located in Kubang Kerian, Kelantan, the majority of the local population is Malay, with limited representation from other ethnic groups. Furthermore, this study only included patients who sought treatment at HPUSM, even though there are nine major hospitals in Kelantan. Patients treated at HPUSM may differ from those seeking care at private or non-tertiary hospitals, where earlier-stage cases are more likely to present. Consequently, the findings may not fully reflect the statewide distribution of breast cancer stages.

For standardisation purposes, the sample size was drawn from patients treated prior to the Covid-19 pandemic, ensuring that the limitations or restrictions imposed during the pandemic did not impact the data. However, as a retrospective study, it is also dependent on existing clinical records, which may be incomplete or subject to documentation biases. The use of patients' residential addresses for spatial mapping also carries a risk of geocoding inaccuracies and recall bias. Additionally, this study did not explore other important factors such as level of education, socioeconomic status, HER2 status, detailed family history of breast cancer, specific genetic information (such as BRCA status), and precise quantification of patient or treatment delays, comorbidities, and healthcare access which could play a role in the progression and outcomes of breast cancer. The absence of socioeconomic status data prevents us from fully accounting for confounding factors; this omission is a likely explanation for the counterintuitive proximity finding, as lower socioeconomic status may correlate with delayed help-seeking despite physical access to care. These variables were not included due to time constraints, but they would provide valuable insights in future research, especially when examining differences between early and late-stage breast cancer patients.

Third, the logistic regression was limited to comparisons between stage III and stage IV breast cancer among late-stage patients, rather than early versus late stages, which restricts the interpretability of factors associated with overall disease advancement. The inclusion of predictors with small subgroup counts (e.g., invasive lobular carcinoma, $n = 6$) resulted in wide confidence intervals and unstable odds ratios. We did not perform post-hoc power analysis or apply penalized regression techniques (such as Firth's correction) due to sample size constraints; these approaches are recommended for future work to improve model stability and precision.

The spatial analysis was performed using Mukim-level aggregation, which may be affected by the Modifiable Areal Unit Problem (MAUP) and edge effects inherent to administrative boundary analyses. Finally, we acknowledge the need for sensitivity analysis regarding the arbitrary 10 km proximity cut-off used in our descriptive analysis. Future studies should test the robustness of this finding using multiple radii (e.g., 5 km and 20 km). Despite these limitations, the integration of clinicopathological and spatial data provides valuable preliminary insight into the geographic and biological determinants of late-stage breast cancer in Kelantan.

In conclusion, this study found a significant clustering of late-stage breast cancer cases and proximity to hospitals is associated with the stage at which breast cancer is diagnosed, with patients living closer to the hospital more likely to present with late-stage disease. Additionally, clinicopathological features specifically progesterone receptor status, were found to be linked to

metastatic disease among late-stage patients, with PR-positive patients showing lower odds of stage IV disease among late-stage patients. These findings highlight the potential of GIS for identifying high-risk areas and the importance of PR status in predicting metastatic disease among late-stage patients. The use of Geographic Information Systems (GIS) can inform targeted early detection strategies such as hotspot-based screening and mobile mammography initiatives in high-risk communities while PR status may serve as a useful marker for risk stratification and personalized management among late-stage patients. Future research integrating HER2 status, socioeconomic factors, and healthcare access variables is warranted to enhance predictive models and guide equitable cancer control strategies.

Acknowledgements

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

Authors declare no conflict of interest.

Author contribution

All authors have contributed to implementation of this research.

Originality Declaration for Figures

All figures included in this manuscript are original and have been created by the authors specifically for the purposes of this study. No previously published or copyrighted images have been used. The authors confirm that all graphical elements, illustrations, and visual materials were generated from the data obtained in the course of this research or designed uniquely for this manuscript.

References

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2024; 74(3)[DOI](#)
2. Azizah AM, Hashimah B, Nirmal K, Siti Zubaidah AR, Puteri NA, Nabihah A, et al. Malaysian National Cancer Registry Report 2012-2016 [https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/2012-2016%20\(MNCRR\)/MNCR_2012-2016_FINAL_\(PUBLISHED_2019\).pdf](https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/2012-2016%20(MNCRR)/MNCR_2012-2016_FINAL_(PUBLISHED_2019).pdf) [23 March 2025].
3. Abdullah NA, Wan Mahiyuddin WR, Muhammad NA, Ali ZM, Ibrahim L, Ibrahim Tamim NS,

- Mustafa AN, Kamaluddin MA. Survival rate of breast cancer patients in Malaysia: a population-based study. *Asian Pacific journal of cancer prevention: APJCP*. 2013; 14(8)[DOI](#)
4. Nordin N, Yaacob NM, Abdullah NH, Mohd Hairon S. Survival Time and Prognostic Factors for Breast Cancer among Women in North-East Peninsular Malaysia. *Asian Pacific journal of cancer prevention: APJCP*. 2018; 19(2)[DOI](#)
 5. Yip CH, Bhoo Pathy N, Teo SH. A review of breast cancer research in malaysia. *The Medical Journal of Malaysia*. 2014; 69 Suppl A
 6. Banerjee S. Spatial Data Analysis. *Annual Review of Public Health*. 2016; 37[DOI](#)
 7. Harinthiran V, W. Z. W. Z, Zakaria AD, Hayati MFM, Ismail AF, Kaur N, Abdullah NH. Geographic Information System (GIS) in Evaluating the Accessibility of Healthcare Facility for Patients with Colon and Rectal Carcinoma in the State of Kelantan and Sabah. *International Journal of Geoinformatics*. 2021; 17(4)
 8. Samat N, Abd Shattar AK, Ghazali S, Bachok NA, Eboy OV, Hasni R, et al. Spatial distribution of breast and cervical cancer incidences in Kelantan: A geographical analysis. *Geografi*. 2013;1:120-31..
 9. Samat N, Jambi D, Musa NS, Shatar AKA, Manan AA, Sulaiman Y. Using a Geographic Information System (GIS) in Evaluating the Accessibility of Health Facilities for Breast Cancer Patients in Penang State, Malaysia. *Kajian Malaysia*. 2010; 28(1)
 10. Shah SA, Neoh H, Rahim SSSA, Azhar ZI, Hassan MR, Safian N, Jamal R. Spatial analysis of colorectal cancer cases in Kuala Lumpur. *Asian Pacific journal of cancer prevention: APJCP*. 2014; 15(3)[DOI](#)
 11. Famarzi S, Kiani B, Hoseinkhani M, Firouraghi N. A gender-specific geodatabase of five cancer types with the highest frequency of occurrence in Iran. *BMC research notes*. 2024; 17(1)[DOI](#)
 12. Firouraghi N, Kiani B, Jafari HT, Learnihan V, Salinas-Perez JA, Raeesi A, Furst M, Salvador-Carulla L, Bagheri N. The role of geographic information system and global positioning system in dementia care and research: a scoping review. *International Journal of Health Geographics*. 2022; 21(1)[DOI](#)
 13. Kiani B, Hashemi Amin F, Bagheri N, Bergquist R, Mohammadi AA, Yousefi M, Faraji H, Roshandel G, Beirami S, Rahimzadeh H, Hoseini B. Association between heavy metals and colon cancer: an ecological study based on geographical information systems in North-Eastern Iran. *BMC cancer*. 2021; 21(1)[DOI](#)
 14. Kiani B, Tabari P, Mohammadi A, Mostafavi SM, Moghadami M, Amini M, Rezaianzadeh A. Spatial epidemiology of skin cancer in Iran: separating sun-exposed and non-sun-exposed parts of the body. *Archives of Public Health = Archives Belges De Sante Publique*. 2022; 80(1)[DOI](#)
 15. Elferink MAG, Pukkala E, Klaase JM, Siesling S. Spatial variation in stage distribution in colorectal cancer in the Netherlands. *European Journal of Cancer (Oxford, England: 1990)*. 2012; 48(8)[DOI](#)
 16. White-Means S, Muruako A. GIS Mapping and Breast Cancer Health Care Access Gaps for African American Women. *International Journal of Environmental Research and Public Health*. 2023; 20(8)[DOI](#)
 17. Madhu B, Srinath KM, Rajendran V, Devi MP, Ashok NC, Balasubramanian S. Spatio-Temporal Pattern of Breast Cancer - Case Study of Southern Karnataka, India. *Journal of clinical and diagnostic research: JCDR*. 2016; 10(4)[DOI](#)
 18. Abdel-Fatah TM, Ball G, Lee AHS, Pinder S, MacMilan RD, Cornford E, Moseley PM. Nottingham Clinico-Pathological Response Index (NPRI) after neoadjuvant chemotherapy (Neo-ACT) accurately predicts clinical outcome in locally advanced breast cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*. 2015; 21(5)[DOI](#)
 19. Tung N, Lin NU, Kidd J, Allen BA, Singh N, Wenstrup RJ, Hartman A, Winer EP, Garber JE. Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. 2016; 34(13)[DOI](#)
 20. Wang R, Zhu Y, Liu X, Liao X, He J, Niu L. The Clinicopathological features and survival

- outcomes of patients with different metastatic sites in stage IV breast cancer. *BMC cancer*. 2019; 19(1)[DOI](#)
21. Sadeghi Yazdankhah S, Javadinia SA, Welsh JS, Mosalaei A. Efficacy of Melatonin in Alleviating Radiotherapy-Induced Fatigue, Anxiety, and Depression in Breast Cancer Patients: A Randomized, Triple-Blind, Placebo-Controlled Trial. *Integrative Cancer Therapies*. 2025; 24[DOI](#)
 22. Shomoossi F, Sheikhmiri S, Chaman R, Zare SS, Welsh JS, Javadinia SA. The Possible Role of Melatonin in Balancing Reactive Oxygen Species (ROS) in Cancer Biology. *Integrative Cancer Therapies*. 2025; 24[DOI](#)
 23. Arifin WN. Sample size calculator (web) [Internet] Available from: <http://wnarifin.github.io> [cited 21 March 2025].
 24. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians*. 2021; 71(3)[DOI](#)
 25. Norsa'adah B, Rahmah MA, Rampal KG, Knight A. Understanding barriers to Malaysian women with breast cancer seeking help. *Asian Pacific journal of cancer prevention: APJCP*. 2012; 13(8)[DOI](#)
 26. Norsa'adah B, Rusli BN, Imran AK, Naing I, Winn T. Risk factors of breast cancer in women in Kelantan, Malaysia. *Singapore Medical Journal*. 2005; 46(12)
 27. Yen SH, Knight A, Krishna M, Muda W, Rufai A. Lifetime Physical Activity and Breast Cancer: a Case-Control Study in Kelantan, Malaysia. *Asian Pacific journal of cancer prevention: APJCP*. 2016; 17(8)
 28. Tesfamariam A, Gebremichael A, Mufunda J. Breast cancer clinicopathological presentation, gravity and challenges in Eritrea, East Africa: management practice in a resource-poor setting. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*. 2013; 103(8)[DOI](#)
 29. Kwak Y, Jang SY, Choi JY, Lee H, Shin DS, Park YH, Kim J, et al. Progesterone Receptor Expression Level Predicts Prognosis of Estrogen Receptor-Positive/HER2-Negative Young Breast Cancer: A Single-Center Prospective Cohort Study. *Cancers*. 2023; 15(13)[DOI](#)
 30. Lu Z, Wang T, Wang L, Ming J. Research progress on estrogen receptor-positive/progesterone receptor-negative breast cancer. *Translational Oncology*. 2025; 56[DOI](#)
 31. Zhang MH, Man HT, Zhao XD, Dong N, Ma SL. Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). *Biomedical Reports*. 2014; 2(1)[DOI](#)