

Duration of Symptoms Predicts Molecular Subtype in Breast Cancer: A Prospective Study from North India Establishing a Clinical Surrogate for Immunohistochemical Classification

Geetika Anand¹, Ashish Kumar Chaudhary¹, Chayanika Kala², Vikash Katiar¹,
Jitendra Kumar Verma³, Prerna Minz¹

^{1,2,4,6}Department of General Surgery, GSVM Medical College, Atal Bihari Vajpayee Medical University, Kanpur, Uttar Pradesh 208002, India

³Department of Pathology, GSVM Medical College, Kanpur, Uttar Pradesh 208002, India

⁵ Department of Radiation Oncology, JK Cancer Institute, Kanpur, Uttar Pradesh, India

Abstract: ***Purpose:** Molecular subtyping of breast cancer by immunohistochemistry (IHC) is often unavailable or unaffordable at the point of first contact in low- and middle-income countries (LMICs). Triple-negative breast cancer (TNBC), which is disproportionately prevalent in Indian populations and concentrated in younger women, carries the worst prognosis and requires urgent therapeutic intervention. We investigated whether duration of symptoms at presentation- a zero-cost clinical parameter- predicts IHC-based molecular subtype and can serve as a surrogate for triaging aggressive disease. **Methods:** This prospective cross-sectional study enrolled 180 consecutive female breast carcinoma patients at a government tertiary centre in North India (September 2024–September 2025). IHC for ER, PR, HER2/Neu, and Ki-67 was performed on formalin-fixed tissue from TruCut biopsy or surgical specimens. Molecular subtypes were assigned per St. Gallen 2013 criteria. Duration of symptoms was correlated with IHC markers using chi-square, Mann–Whitney U, and Spearman tests. ROC analysis evaluated Ki-67 as a predictor of ER negativity. Kaplan–Meier and life table analyses assessed event-free survival (EFS). **Results:** Duration showed the strongest correlation with molecular subtype ($\chi^2=224.56$, $p<0.001$): 82.5% of patients presenting within ≤ 3 months were TNBC, while 100% of those presenting after >9 months were Luminal A. Duration inversely correlated with Ki-67 (Spearman $\rho=-0.840$, $p<0.001$). Mean duration was 8.1 months for ER-positive versus 2.8 months for ER-negative tumours ($p<0.001$). Ki-67 predicted ER negativity with AUC=0.964 (cutoff 55%). TNBC accounted for 21.7% of subtypes- double the Western prevalence- and was concentrated in patients aged ≤ 40 (50.0% of TNBC). Six-month EFS was 48.4% for TNBC versus 100% for Luminal A (log-rank $p<0.001$). All deaths occurred in the TNBC subgroup. **Conclusion:** Symptom duration is a powerful clinical correlate of molecular subtype, driven by subtype-specific tumour proliferative kinetics. In settings where IHC is inaccessible or unaffordable, short symptom duration (≤ 3 months) serves as a zero-cost clinical flag for aggressive subtypes requiring urgent intervention.*

Keywords: Breast cancer, Duration of symptoms, Molecular subtypes, Ki-67, Triple-negative breast cancer, Tumour growth kinetics, India, Resource-limited settings

1. Introduction

Breast carcinoma is a biologically heterogeneous disease comprising molecular subtypes- Luminal A, Luminal B, HER2-enriched, and triple-negative (TNBC)- that carry fundamentally different growth kinetics, therapeutic vulnerabilities, and clinical outcomes [1, 2]. The immunohistochemical (IHC) assessment of estrogen receptor (ER), progesterone receptor (PR), HER2/Neu, and Ki-67 proliferation index serves as the standard surrogate for gene expression-based classification and is indispensable for treatment planning [3, 4]. However, IHC panels remain inaccessible or unaffordable for a substantial proportion of breast cancer patients in low- and middle-income countries (LMICs). In India, the cost of a four-marker IHC panel (approximately ₹3,000–6,000 / US\$35–70) often exceeds the daily household income for patients presenting to government tertiary centres, and many district-level hospitals lack the infrastructure for IHC entirely [5].

India now bears the world's heaviest absolute burden of breast cancer mortality, with approximately 192,020 new cases and

98,337 deaths annually- the highest death toll from breast cancer of any country worldwide [6, 7]. Indian breast cancer diverges profoundly from Western populations in three critical ways. First, patients present a decade younger, with peak incidence in the 41–50-year age group rather than the 60–70-year peak in the West [8, 9]. Second, the proportion of TNBC is disproportionately high- 21–31% in Indian cohorts versus 10–17% in Western populations- and is rising, particularly in younger women below 40 years, even at tertiary care centres where patients typically present at advanced stages rather than through screening [10–12]. This concentration of the most aggressive subtype in the youngest patients, in a country where late presentation is the norm and IHC access is limited, represents a uniquely dangerous epidemiological convergence. Third, the majority of patients present with locally advanced disease rather than screen-detected early-stage cancers, meaning that the disease burden at a typical Indian government hospital is biologically more aggressive from the outset [8, 13].

The biological basis for expecting symptom duration to predict molecular subtype is compelling. Tumour doubling times differ dramatically by subtype: TNBC doubles in

approximately 103 ± 43 days versus 241 ± 166 days for Luminal A- a 2.3-fold kinetic divergence [14]. This disparity originates in fundamental molecular machinery. TNBC arises predominantly from undifferentiated basal/myoepithelial progenitor cells characterised by TP53 mutations (>80%), BRCA1 dysfunction, and PI3K/AKT pathway hyperactivation, all converging on unrestricted mitotic activity [15, 16]. Critically, the loss of ER function removes the ER-mediated transcriptional programmes that activate cell cycle inhibitors (p21, p27) in luminal tumours, eliminating the proliferative brake [17, 18]. The phenotypic result is a tumour that divides faster, recruits angiogenesis earlier, invades lymphovascular spaces sooner, and generates symptomatic masses within weeks to months. Luminal A tumours, constrained by hormone receptor signalling, grow slowly enough to remain subclinical for months to years.

Interval cancer studies in screening populations have indirectly validated this kinetic divergence: tumours detected symptomatically between screening rounds are enriched for TNBC, high grade, and genomic instability [19, 20]. Yet the direct clinical corollary- that self-reported symptom duration quantitatively predicts molecular subtype- has been studied in only two populations globally: a South African cohort [21] and a Saudi Arabian analysis [22]. No data exist from India, despite its uniquely high TNBC burden and uniquely constrained IHC access. A zero-cost clinical parameter that predicts likely molecular subtype at first contact could transform triage pathways in the Indian healthcare system and other LMICs.

We hypothesised that duration of symptoms at presentation is strongly correlated with IHC-based molecular subtype classification, driven by subtype-specific tumour proliferative kinetics indexed by Ki-67. We additionally evaluated Ki-67 as a continuous predictor of ER negativity and assessed short-term event-free survival by molecular subtype.

2. Materials and methods

Study design and population

This prospective cross-sectional study was conducted in the Department of General Surgery, GSVM Medical College (a government tertiary care teaching hospital), Kanpur, India, from September 2024 to September 2025. The study was approved by the Institutional Ethics Committee (IEC No. EC/401/September/2024, dated 08 October 2024) and written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

A total of 180 consecutive female patients with histopathologically confirmed primary invasive breast carcinoma were enrolled. Inclusion required availability of adequate tumour tissue for IHC analysis of all four markers (ER, PR, HER2/Neu, Ki-67) from either TruCut biopsy or surgical specimen, and complete clinical records including demographic data, clinical examination findings, imaging reports, and histopathological reporting with TNM staging. Exclusion criteria were: (a) receipt of neoadjuvant chemotherapy prior to tissue sampling, as treatment can significantly alter receptor expression and tumour morphology, confounding the assessment of native tumour

biology; (b) male breast carcinoma, due to its distinct epidemiological and biological profile; (c) recurrent breast carcinoma without fresh tissue sampling from the recurrent lesion; and (d) incomplete IHC data where one or more of the four markers could not be evaluated.

Clinical assessment

Duration of symptoms was defined as the interval between the patient's first self-recognised breast symptom (most commonly a palpable lump) and the date of histopathological diagnosis, recorded in months from clinical history. Clinico-pathological parameters recorded on a structured proforma included age, menstrual status, tumour size (maximum dimension), histological type (WHO Classification of Tumours, 5th Edition), histological grade (Nottingham/Bloom–Richardson modification), axillary lymph node status, lymphovascular invasion (LVI), perineural invasion (PNI), skin/nipple involvement, and clinical stage (AJCC 8th Edition).

Immunohistochemistry- technique and scoring

IHC was performed on 4- μ m FFPE tissue sections using standardised avidin-biotin-peroxidase methodology with heat-induced epitope retrieval (HIER) in citrate buffer (pH 6.0). Primary antibodies included ER (clone 1D5), PR (clone PgR636), HER2/Neu (polyclonal), and Ki-67 (clone MIB-1). ER and PR were scored using the Allred method (proportion score 0–5 plus intensity score 0–3; total 0–8; scores ≥ 3 positive; $\geq 1\%$ nuclear staining defining positivity per ASCO/CAP 2020) [23]. HER2/Neu was scored per ASCO/CAP 2018 criteria (0/1+ negative, 2+ equivocal, 3+ positive) [24]. Ki-67 was quantified as the percentage of positively staining tumour nuclei across hotspot areas; a 20% cutoff distinguished high from low proliferation (St. Gallen 2013) [4]. Evaluable-case denominators were used: ER (n=176), PR (n=175), HER2 (n=174), Ki-67 (n=171), molecular subtype (n=175). Subtypes were assigned as: Luminal A (ER/PR+, HER2–, Ki-67 <20%), Luminal B (ER/PR+, HER2– with Ki-67 $\geq 20\%$, or ER/PR+ with HER2+), HER2-enriched (ER–, PR–, HER2+), and TNBC (ER–, PR–, HER2–).

Statistical analysis

Chi-square and Fisher's exact tests assessed categorical associations. Mann–Whitney U test compared duration between IHC-positive and -negative groups. Spearman's rank correlation (ρ) quantified monotonic relationships. ROC curve analysis with Youden index determined optimal Ki-67 cutoffs for predicting ER negativity, HER2 positivity, and lymph node metastasis. Kaplan–Meier analysis with log-rank test assessed event-free survival (EFS; event = recurrence or death) on 158 patients with known follow-up (median 6 months). Actuarial life tables used 2-month intervals. All tests were two-sided; $p < 0.05$ was significant. Analysis was performed using Python 3.11 (SciPy 1.11 [25], lifelines 0.27 [26], scikit-learn 1.3 [27]). This study is reported per the STROBE statement [28] and REMARK guidelines [29].

3. Results

Patient demographics and tumour characteristics

Mean age was 48.9 ± 11.6 years (median 47, range 25–90), with peak incidence in the 41–50-year group (36.7%). Pre-menopausal women constituted 53.3%. Mean tumour size was 4.16 ± 1.97 cm; 81.7% were T2 tumours. IDC (NOS) was the predominant histological type (82.2%). Grade II was the most common (59.8%), followed by Grade III (31.8%). Lymph node involvement was present in 68.3%, LVI in 64.4%, PNI in 23.9%. Stage IIB was the commonest presentation (43.3%); only 1.1% presented at Stage I. Mean duration of symptoms was 6.4 ± 3.3 months (median 6, range 1–18).

ER positivity was 68.2% (120/176), PR 65.7% (115/175), HER2 overexpression 17.2% (30/174), and high Ki-67 ($\geq 20\%$) 60.8% (104/171). ER–PR concordance was 96.6%. Luminal A was the commonest subtype at 38.3% (67/175), followed by Luminal B 29.7% (52/175), TNBC 21.7% (38/175), and HER2-enriched 10.3% (18/175).

TNBC: disproportionate prevalence in young Indian women

The TNBC prevalence of 21.7% in our cohort is approximately double the 10–17% reported in Western populations, consistent with Indian pooled estimates of 27–31% [10, 11]. Strikingly, TNBC was concentrated in younger patients: 66.7% of patients aged 21–30 years and 39.5% of those aged 31–40 years were TNBC, compared with 0% in patients >70 years ($p < 0.001$). Overall, 50.0% (19/38) of all TNBC patients were aged ≤ 40 years. This concentration of the most aggressive subtype in the youngest demographic, at a government tertiary centre where patients present predominantly with symptomatic rather than screen-detected disease, underscores a critical public health challenge: the very patients who are most likely to have aggressive biology are the least likely to access timely IHC-guided treatment in the Indian healthcare system.

Duration of symptoms predicts molecular subtype

Duration demonstrated the strongest single statistical association with molecular subtype observed in the study ($\chi^2 = 224.56$, $df = 9$, $p < 0.001$; Table 2). The relationship followed a stepwise gradient: among patients presenting within ≤ 3 months, 82.5% were TNBC and 17.5% were HER2-enriched, with zero Luminal A or Luminal B cases. At 4–6 months, Luminal B predominated (66.0%). At 7–9 months, Luminal A became the dominant subtype (66.0%). At >9 months, 100% were Luminal A.

Duration showed highly significant associations with all four IHC markers (all $p < 0.001$; Table 2). ER positivity rose from 0% in the ≤ 3 -month group to 100% in the >9 -month group ($\chi^2 = 124.10$). Ki-67 was high in 100% of patients with ≤ 3 -month duration versus 0% of those with >9 -month duration ($\chi^2 = 108.89$). Mean duration was 8.1 ± 2.6 months for ER-positive versus 2.8 ± 1.1 months for ER-negative tumours (Mann–Whitney $p < 0.001$), and 4.4 ± 1.9 months for Ki-67-high versus 9.5 ± 2.4 months for Ki-67-low tumours ($p < 0.001$).

Spearman correlation between duration and Ki-67 was $\rho = -0.840$ ($p < 0.001$)—among the strongest reported for any clinical-biological correlation in breast cancer. Duration also correlated with age ($\rho = +0.439$, $p < 0.001$) and inversely with tumour size ($\rho = -0.352$, $p < 0.001$). The duration–subtype association persisted across invasion markers: PNI was present in 70.0% of ≤ 3 -month patients but 0% of >9 -month patients ($p < 0.001$); LVI showed a parallel gradient (90.0% versus 21.9%; $p < 0.001$).

The biological mechanism: proliferative kinetics govern symptom onset

The data support Ki-67 as the primary biological mediator of the duration–subtype relationship. Grade-specific Ki-67 distributions illustrate the mechanism: Grade I tumours had mean Ki-67 of 12.6% versus 69.4% in Grade III (Spearman $\rho = 0.778$, $p < 0.001$). Grade III tumours were 94.4% ER-negative, 98.1% Ki-67-high, and dominated the ≤ 3 -month category (94.9%). This creates a coherent mechanistic cascade: high Ki-67 $\rightarrow$ rapid tumour volume doubling (median ~ 103 days for TNBC [14]) $\rightarrow$ early lymphovascular and perineural invasion $\rightarrow$ early symptom generation $\rightarrow$ short presentation duration. The inverse chain operates for indolent Luminal A tumours: low Ki-67 $\rightarrow$ slow doubling (~ 241 days [14]) $\rightarrow$ prolonged subclinical growth $\rightarrow$ long symptom duration.

Critically, the duration–tumour size correlation was only moderate ($\rho = -0.352$), confirming that short-duration tumours are not simply larger tumours detected earlier, but are biologically different tumours that grow faster, invade earlier, and generate symptoms sooner at any given size. This distinction—between delay-driven stage progression and biology-driven rapid presentation—is the central insight of the study.

ROC analysis further quantified this relationship: Ki-67 as a continuous variable predicted ER negativity with AUC=0.964 (95% CI 0.935–0.993; optimal cutoff 55%, sensitivity 96.3%, specificity 97.4%). Ki-67 also showed good performance for lymph node metastasis (AUC=0.809, cutoff 18%) and acceptable performance for HER2 positivity (AUC=0.751, cutoff 30%) (Table 3).

Survival analysis

Kaplan–Meier analysis ($n = 158$) revealed a pronounced subtype gradient in EFS (log-rank $\chi^2 = 54.42$, $p < 0.001$; Table 4). Six-month EFS was 100% for Luminal A, 96.0% for Luminal B, 66.7% for HER2-enriched, and 48.4% for TNBC. ER-positive EFS was 98.1% versus 54.4% for ER-negative ($\chi^2 = 49.77$, $p < 0.001$). All 7 deaths occurred in the TNBC subgroup. Life table analysis showed TNBC events beginning in the first 2-month interval ($q = 0.032$) and accelerating to $q = 0.615$ in the 6–8-month interval (cumulative survival 0.335), whereas Luminal B and HER2-enriched events were confined to the 6–8-month interval. All recurrences and deaths occurred in patients with ≤ 6 -month symptom duration.

4. Discussion

Principal findings

The central finding is that duration of symptoms at presentation is the strongest single predictor of molecular

subtype ($\chi^2=224.56$) in this prospective cohort- exceeding the chi-square values for all other clinico-pathological-IHC associations by a wide margin. The near-complete subtype segregation by duration (82.5% TNBC at ≤ 3 months, 100% Luminal A at >9 months) has not been reported with this magnitude from any population globally.

Tumour biology explains the paradox

Our findings embody the “delay paradox” in breast cancer: patients who present earliest have the worst biology and worst outcomes [30, 31]. The biological resolution is that shorter duration does not cause worse biology- rather, worse biology causes shorter duration by accelerating tumour growth and symptom onset. The mechanistic chain has three pillars.

First, proliferative rate directly governs symptom timing. The Spearman ρ of -0.840 between duration and Ki-67 indicates that approximately 70% of the variance in symptom duration is explained by proliferative rate. TNBC’s molecular features- TP53 loss ($>80\%$), BRCA1 dysfunction, PI3K/AKT hyperactivation, and absence of ER-mediated cell cycle restraint- collectively produce doubling times 2.3-fold faster than Luminal A [14, 15]. A TNBC starting from a single transformed cell reaches palpable size (~ 1 cm) in approximately 6–8 months of exponential growth; a Luminal A tumour requires 14–18 months- timelines consistent with our mean durations (ER-negative 2.8, ER-positive 8.1 months).

Second, the molecular machinery governing proliferation simultaneously determines receptor status. ER-mediated transcription activates differentiation and cell cycle arrest programmes [17, 18]. Loss of ER removes both the therapeutic target and the proliferative brake. This is not coincidence but direct mechanistic linkage: the same molecular event (ER loss) that makes a tumour receptor-negative also makes it fast-growing, explaining why duration and receptor status are so tightly correlated.

Third, rapid proliferation has downstream invasion consequences. LVI was 90.0% in ≤ 3 -month patients versus 21.9% in >9 -month patients; PNI was 70.0% versus 0%. These are not merely consequences of tumour size- duration’s correlation with tumour size was only moderate ($\rho=-0.352$)- but of accelerated tumour biology driving earlier invasion regardless of size.

The Indian TNBC burden: a crisis in need of clinical tools

Our TNBC prevalence of 21.7% is approximately double the 10–17% in Western populations, consistent with the Indian pooled estimate of 27–31% from meta-analyses [10, 11]. The concentration of TNBC in women aged ≤ 40 (50.0% of our TNBC patients) at a government tertiary centre- where patients present predominantly with symptomatic disease rather than through screening- has three important implications. First, it confirms that the rising incidence of aggressive breast cancer subtypes in younger Indian women is not an artifact of screening bias; these are symptomatic cancers presenting to institutions where mammographic screening is not routinely performed. Second, it highlights that the patients most in need of urgent molecular characterisation (young women with rapidly growing TNBC) are precisely those for whom IHC is least likely to be

available at first contact in the Indian healthcare system, where the four-marker IHC panel costs ₹3,000–6,000 and is unavailable at many district hospitals. Third, it creates a clinical imperative for surrogate markers that can identify these patients at zero cost.

Duration of symptoms meets this need. In our data, a patient presenting to a primary or district hospital with a rapidly growing breast lump of ≤ 3 months has an $\geq 80\%$ probability of harbouring TNBC and should be immediately referred for biopsy and IHC workup at a tertiary centre, even before molecular results are available. This triage principle costs nothing, requires no laboratory infrastructure, and can be implemented at any level of the healthcare system.

Comparison with existing literature

Only two studies globally have directly correlated symptom duration with molecular subtype. The Eastern Cape, South Africa study (2025) found significant duration–subtype associations (Fisher’s exact $p<0.001$) with TNBC and HER2-enriched presenting earlier [21]; our $\chi^2=224.56$ substantially exceeds their result. Alsalamah (2025, Saudi Arabia) identified subtype-specific delay thresholds (38 days for TNBC, 85 days for ER+/PR+/HER2-) using Markov modelling [22], consistent with our TNBC concentration in the ≤ 3 -month group. No prior study from India has examined this relationship, and no study globally has proposed using duration as a practical clinical surrogate marker for molecular subtype at first contact.

Strengths and limitations

Strengths include the prospective design with consecutive enrolment, comprehensive IHC panel with molecular subtyping, systematic duration recording, evaluable-case denominators with transparent exclusion, multiple complementary statistical methods, and STROBE/REMARK compliance. The study setting- a government tertiary centre in North India, not a private hospital- ensures that the results are directly relevant to the patient population for whom IHC accessibility is most constrained.

Limitations include the short follow-up (6 months), single-centre design, absence of FISH for HER2-equivocal cases ($n=3$), self-reported symptom duration (subject to recall bias), and lack of formal multiple comparison correction across 52 tests- though the predominance of $p<0.001$ results mitigates Type I error concerns. External validation in a multi-centre Indian cohort with longer follow-up is needed.

5. Conclusion

Duration of symptoms at presentation is a powerful clinical correlate of IHC-based molecular subtype in breast carcinoma, driven by subtype-specific tumour proliferative kinetics. The near-complete subtype segregation by duration interval ($\chi^2=224.56$)- with TNBC dominating short-duration and Luminal A dominating long-duration presentations- positions symptom duration as a clinically accessible, zero-cost surrogate marker. This has particular relevance for India and other LMICs, where IHC access is limited, TNBC prevalence is disproportionately high, patients present with symptomatic rather than screen-detected disease, and the

youngest patients harbour the most aggressive biology. Multi-centre validation is recommended.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interest: The authors have no relevant financial or non-financial interests to disclose.

Ethics approval: This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Institutional Ethics Committee, GSVM Medical College, Kanpur (IEC No. EC/401/September/2024, dated 08 October 2024).

Consent to participate: Informed consent was obtained from all individual participants included in the study.

Consent for publication: Not applicable (no individually identifying data).

Author contributions: GA: Data collection, statistical analysis, manuscript preparation. AKC: Study conception, supervision, critical review (Guide). CK: IHC interpretation, pathological review. VK: Surgical management, data review. JKV: Radiation oncology input, review. PM: Critical review. All authors read and approved the final manuscript.

Availability of data: The datasets analysed during the current study are available from the corresponding author on reasonable request.

References

- [1] Perou CM, Sørbye T, Eisen MB et al (2000) Molecular portraits of human breast tumours. *Nature* 406:747–752
- [2] Harbeck N, Penault-Llorca F, Cortes J et al (2019) Breast cancer. *Nat Rev Dis Primers* 5:66
- [3] Goldhirsch A, Winer EP, Coates AS et al (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus 2013. *Ann Oncol* 24:2206–2223
- [4] Dowsett M, Nielsen TO, A'Hern R et al (2011) Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer Working Group. *J Natl Cancer Inst* 103:1656–1664
- [5] Prajapati A, Gupta S, Mistry B, Gupta S (2025) Comprehensive overview of breast cancer in India. *JCO Glob Oncol*. <https://doi.org/10.1200/GO-25-00083>
- [6] Bray F, Laversanne M, Sung H et al (2024) Global cancer statistics 2022: GLOBOCAN estimates. *CA Cancer J Clin* 74:229–263
- [7] Sathishkumar K, Chaturvedi M, Das P et al (2022) Cancer incidence estimates for 2022 & projection for 2025: NCRP, India. *Indian J Med Res* 156:598–607
- [8] Malvia S, Bagadi SA, Dubey US, Saxena S (2017) Epidemiology of breast cancer in Indian women. *Asia Pac J Clin Oncol* 13:289–295
- [9] Bisht M, Agarwal S, Agrawal R (2018) Clinicopathological and immunohistological profile of IDC breast in western UP. *Indian J Pathol Oncol* 5:619–624
- [10] Jain A, Bhayana R, Vlasschaert C et al (2021) Prevalence of molecular subtypes of breast cancer in India: a systematic review and meta-analysis. *Cancer Res Stat Treat* 4:252–260
- [11] Kulkarni A, Stroup AM, Paddock LE et al (2020) Breast cancer incidence and mortality by molecular subtype. *Cancer Health Disparities* 4:e1–e17
- [12] Sharma M, Sharma JD, Sarma A et al (2014) Triple negative breast cancer in people of north east India. *Asian Pac J Cancer Prev* 15:4507–4511
- [13] Doval DC, Radhakrishna S, Tripathi R et al (2020) A multi-institutional real world data study from India of 3453 non-metastatic breast cancer patients. *Sci Rep* 10:5886
- [14] Dahan M, Bhatt AA, Taurin S (2021) Has tumour doubling time in breast cancer changed over the past 80 years? A systematic review. *Cancer Med* 10:5203–5217
- [15] Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490:61–70
- [16] Bianchini G, De Angelis C, Licata L, Gianni L (2022) Treatment landscape of triple-negative breast cancer. *Nat Rev Clin Oncol* 19:91–113
- [17] Musgrove EA, Sutherland RL (2009) Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer* 9:631–643
- [18] Sørlie T, Perou CM, Tibshirani R et al (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses. *Proc Natl Acad Sci USA* 98:10869–10874
- [19] Pal Choudhury P, Hoover RN, Engel C et al (2024) Genetic landscape of interval and screen detected breast cancer. *npj Precis Oncol* 8:131
- [20] Domingo L, Blanch J, Servitja S et al (2013) Biological characteristics of interval cancers: a role for biomarkers in breast cancer screening. *Breast Cancer Res* 15:R82
- [21] [Eastern Cape South Africa study] (2025) Duration of symptoms and molecular subtypes in breast cancer. *PLOS ONE* [citation to be confirmed]
- [22] Alsalamah YA (2025) Critical diagnostic delay thresholds in breast cancer: a molecular subtype-based causal analysis from Saudi Arabia. *Cureus* 17:e76058
- [23] Allison KH, Hammond MEH, Dowsett M et al (2020) ER and PR testing in breast cancer: ASCO/CAP guideline update. *J Clin Oncol* 38:1346–1366
- [24] Wolff AC, Hammond MEH, Allison KH et al (2018) HER2 testing in breast cancer: ASCO/CAP guideline focused update. *J Clin Oncol* 36:2105–2122
- [25] Virtanen P, Gommers R, Oliphant TE et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17:261–272
- [26] Davidson-Pilon C (2019) lifelines: survival analysis in Python. *J Open Source Softw* 4:1317
- [27] Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- [28] von Elm E, Altman DG, Egger M et al (2007) The STROBE statement: guidelines for reporting observational studies. *Lancet* 370:1453–1457

- [29] McShane LM, Altman DG, Sauerbrei W et al (2005) REMARK guidelines for tumour marker prognostic studies. *J Natl Cancer Inst* 97:1180–1184
- [30] Richards MA, Westcombe AM, Love SB et al (1999) Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet* 353:1119–1126
- [31] Caplan L (2014) Delay in breast cancer: implications for stage at diagnosis and survival. *Front Public Health* 2:87
- [32] Unger-Saldaña K (2014) Challenges to the early diagnosis and treatment of breast cancer in developing countries. *World J Clin Oncol* 5:465–477
- [33] Nielsen TO, Leung SCY, Rimm DL et al (2021) Assessment of Ki67 in breast cancer: updated recommendations from the IKWG. *J Natl Cancer Inst* 113:808–819
- [34] Barodawala SM, Jain SA, Panchal S et al (2024) Clinicopathological correlation of ER, PR, HER2 in 12,808 breast carcinoma patients. *Breast Cancer (Auckl)* 18:11782234241273543
- [35] Rhodes A, Jasani B, Balaton AJ et al (2000) Frequency of oestrogen and progesterone receptor positivity in 7,016 breast carcinomas. *J Clin Pathol* 53:688–696
- [36] Effi AB, Aman NA, Kouli BS et al (2017) Immunohistochemical determination of ER and PR in breast cancer: 302 patients. *BMC Cancer* 17:115
- [37] Agodirin O, Olatoke S, Rahman G et al (2021) Determinants of delayed presentation and advanced-stage diagnosis of breast cancer in Africa: a systematic review. *Asian Pac J Cancer Prev* 22:1007–1017
- [38] Ranganathan P, Aghoram R, Nair SM (2024) Statistical tools for data collection, management, and analysis. *Perspect Clin Res* 15:209–212

Tables

Table 1: Consolidated clinico-pathological and IHC profile (n=180)

Parameter	Category	n	%
Age group	41–50 (peak)	66	36.7
	Mean±SD	48.9±11.6	
Duration	≤3 months	40	22.2
	4–6 months	57	31.7
	7–9 months	51	28.3
	>9 months	32	17.8
Menstrual	Pre-menopausal	96	53.3
Tumour size	T2 (2.1–5 cm)	147	81.7
Histological type	IDC (NOS)	148	82.2
Grade (n=179)	I / II / III	15/107/57	8.4/59.8/31.8
LN involvement	Positive	123	68.3
LVI	Present	116	64.4
PNI	Present	43	23.9
Stage	IIB (commonest)	78	43.3
ER (n=176)	Positive	120	68.2
PR (n=175)	Positive	115	65.7
HER2 (n=174)	Positive	30	17.2
Ki-67 (n=171)	High (≥20%)	104	60.8
Subtype (n=175)	Luminal A / B / HER2-e / TNBC	67/52/18/38	38.3/29.7/10.3/21.7

Table 2: Duration of symptoms versus molecular subtype and IHC markers

Duration	n	Lum A	Lum B	HER2-e	TNBC	ER+ %	Ki67 Hi%
≤3 mo	40	0 (0%)	0 (0%)	7 (17.5%)	33 (82.5%)	0	100
4–6 mo	53	2 (3.8%)	35 (66.0%)	11 (20.8%)	5 (9.4%)	70.4	92.5
7–9 mo	50	33 (66.0%)	17 (34.0%)	0 (0%)	0 (0%)	100	34.7
>9 mo	32	32 (100%)	0 (0%)	0 (0%)	0 (0%)	100	0
p-value		$\chi^2=224.56$	$p<0.001$			$\chi^2=124.10$	$\chi^2=108.89$

Table 3: ROC curve analysis- Ki-67 as predictor

Outcome	AUC (95% CI)	Cutoff	Sensitivity	Specificity
ER negativity	0.964 (0.935–0.993)	55%	96.30%	97.40%
HER2 positivity	0.751 (0.663–0.839)	30%	96.70%	71.70%
LN metastasis	0.809 (0.744–0.874)	18%	87.10%	69.10%

Table 4: Kaplan–Meier event-free survival by molecular subtype (n=158)

Subtype	n	Events	6-mo EFS	Log-rank vs TNBC	Deaths
Luminal A	59	0	100.00%	$p<0.001$	0
Luminal B	50	2	96.00%	$p<0.001$	0
HER2-enriched	15	5	66.70%	$p=0.182$ (NS)	0
TNBC	32	16	48.40%	(Reference)	7

Overall log-rank: $\chi^2=54.42$, $df=3$, $p<0.001$. ER+ EFS=98.1% vs ER– 54.4% ($p<0.001$). All 7 deaths in TNBC subgroup.

Figure legends

Figure 1: 100% stacked bar chart showing molecular subtype distribution by duration of symptoms. TNBC dominates ≤3-month presentations (82.5%) while Luminal A constitutes 100% of >9-month presentations ($\chi^2=224.56$, $p<0.001$).

Figure 2: ROC curves for Ki-67 predicting (A) ER negativity (AUC=0.964), (B) HER2 positivity (AUC=0.751), (C) LN metastasis (AUC=0.809). Red dots indicate optimal cutoff points determined by Youden index.

Figure 3: Kaplan–Meier event-free survival curves by molecular subtype (n=158). TNBC 6-month EFS=48.4%; Luminal A=100%. Overall log-rank $p<0.001$.