**Breast Cancer Recurrence Analysis Report**

**Introduction**

Breast cancer recurrence remains an important clinical challenge, with implications for survival, treatment planning, and long-term quality of life. Identifying patient- and tumor-level predictors of recurrence can help clinicians stratify risk and improve individualized treatment strategies. Epidemiologic evidence highlights age, tumor receptor status, and stage as established risk factors, while disparities by race and socioeconomic characteristics have also been observed. In this study, we used data from a large hospital-based cancer registry to investigate predictors of first recurrence in breast cancer patients. Given that patients were treated across multiple healthcare facilities, we applied multilevel statistical modeling to account for clustering at the facility level.

**Methods**

**Data Source and Sample:** Data were derived from a multi-site hospital cancer registry encompassing breast cancer cases diagnosed and treated between 2018 and 2023. After deduplication by unique medical record number (MRN), the analytic dataset consisted of 3,257 patients, of whom 2,715 did not experience recurrence and 542 experienced a first recurrence during follow-up.

**Outcome:** The primary outcome was first recurrence, defined using the variable “*Type. 1st.Recurrence*”. Patients were classified as No Recurrence (coded 0) if no recurrence was documented, or Recurrence (coded 1) if any type of recurrence was reported.

**Covariates:** The primary predictors of interest included age category, race category, estrogen receptor (ER) status, and progesterone receptor (PR) status. Age at diagnosis was categorized into four groups to reflect both clinical relevance and the observed distribution of patients: 0–40 years, 40–60 years, 60–80 years, and ≥80 years, with the youngest group (0–40 years) serving as the reference category. Race was harmonized into three categories, with Black patients designated as the reference group; White patients were retained as a separate category, while patients who identified as Asian, Native American, Pacific Islander, or multiracial were collapsed into an “Other” category to address sparse data issues. ER and PR statuses were each coded as binary variables, with tumors classified as either positive or negative. To account for potential clustering of patients within healthcare facilities, a random intercept for facility was incorporated into the model to capture between-facility variation in recurrence risk.

**Data Management:** Data preparation involved several steps. First, multiple records for the same patient were identified using the medical record number (MRN), and only the earliest record by diagnosis date was retained to ensure that each patient contributed a single observation. Finally, a complete-case approach was applied, whereby observations with missing values in the outcome or any of the key predictors were excluded from the analytic dataset.

**Statistical Analysis:** Patient characteristics were first summarized according to recurrence status using descriptive statistics. Table 1 presents distributions of sociodemographic and clinical variables stratified by recurrence. Categorical variables, including race, ER status, PR status, HER2 status, cancer stage, insurance type, and facility of treatment, were compared using chi-squared tests while age employed two sample t-test. To preliminarily assess the strength of association between each predictor and recurrence, bivariate logistic regression models were fitted. Odds ratios (ORs) with 95% confidence intervals (CIs) were estimated for each variable individually.

Subsequently, a generalized linear mixed-effects model (GLMM) with a logit link was applied to simultaneously evaluate the effects of age category, race category, ER status, and PR status on recurrence risk, while accounting for clustering of patients within facilities. A random intercept for facility was included to capture heterogeneity in baseline recurrence risk across treatment sites. Odds ratios with 95% CIs were reported for all fixed effects. Model performance was assessed by examining the Akaike Information Criterion (AIC), the intraclass correlation coefficient (ICC) derived from the variance of the facility-level random effect, and marginal and conditional R² to quantify explained variance. Discriminatory capacity was evaluated by calculating the area under the receiver operating characteristic curve (AUC) for both marginal predictions (fixed effects only) and conditional predictions (fixed plus random effects)

**Patient characteristics by recurrence status**

The mean age was similar between patients without recurrence (62.7 years, SD 12.7) and those with recurrence (63.1 years, SD 15.9), and no significant difference was observed. However, when age was analyzed categorically, patients ≤40 years represented a larger proportion of those with recurrence (9.8%) compared with those without recurrence (3.8%), while patients aged 60–80 years were more common in the non-recurrence group (53.8% vs. 41.1%). Significant proportional differences were also observed by race: White patients comprised most of the cohort, but a greater proportion of Black patients experienced recurrence (6.8%) compared to non-recurrence (3.1%).

Patients with HER2-positive disease accounted for 19.1% of those with recurrence compared to 13.7% of those without, while ER negativity was nearly twice as common among patients with recurrence (23.6% vs. 12.4%). Stage at diagnosis demonstrated a striking gradient: nearly one quarter (24.5%) of patients with recurrence presented with stage IV disease, compared to just 0.4% of those without recurrence, whereas stage I disease dominated the non-recurrence group (80.6% vs. 41.1%). Patients insured by Medicaid/Medicare or classified under “Other” insurance were more frequently represented in the recurrence group, while those with private insurance were more common in the non-recurrence group. Facility-level differences were apparent, with some sites, such as St. Francis and Bergan, contributing disproportionately to the recurrence group.

Table 1. Patient characteristics by recurrence status

| **Characteristic** | **No Recurrence (n = 2,715)** | **Recurrence (n = 542)** | **p-value** |
| --- | --- | --- | --- |
| **Age, mean (SD)** | 62.66 (12.70) | 63.09 (15.89) | 0.492 |
| **Race** |  |  | **<0.001** |
| – Black | 84 (3.1%) | 37 (6.8%) |  |
| – Other | 73 (2.7%) | 20 (3.7%) |  |
| – White | 2,558 (94.2%) | 485 (89.5%) |  |
| **HER2 status** |  |  | **0.002** |
| – Negative | 2,291 (86.3%) | 429 (80.9%) |  |
| – Positive | 365 (13.7%) | 101 (19.1%) |  |
| **ER status** |  |  | **<0.001** |
| – Negative | 330 (12.4%) | 125 (23.6%) |  |
| – Positive | 2,326 (87.6%) | 405 (76.4%) |  |
| **Insurance** |  |  | **<0.001** |
| – Medicaid | 77 (2.8%) | 24 (4.4%) |  |
| – Medicare | 1,264 (46.6%) | 246 (45.4%) |  |
| – Private | 1,085 (40.0%) | 173 (31.9%) |  |
| – Other | 289 (10.6%) | 99 (18.3%) |  |
| **Cancer stage** |  |  | **<0.001** |
| – Stage 0 | 42 (1.5%) | 0 (0.0%) |  |
| – Stage I | 2,187 (80.6%) | 219 (40.4%) |  |
| – Stage II | 319 (11.7%) | 93 (17.2%) |  |
| – Stage III | 155 (5.7%) | 97 (17.9%) |  |
| – Stage IV | 12 (0.4%) | 133 (24.5%) |  |
| **Age categories** |  |  | **<0.001** |
| – 0 to 40 years | 104 (3.8%) | 53 (9.8%) |  |
| – 40 to 60 years | 935 (34.4%) | 172 (31.7%) |  |
| – 60 to 80 years | 1,460 (53.8%) | 223 (41.1%) |  |
| – Over 80 years | 216 (8.0%) | 94 (17.3%) |  |
| **Campus** |  |  | **<0.001** |
| – Other/Unknown | 20 (0.7%) | 4 (0.7%) |  |
| – Bergan | 262 (9.7%) | 97 (17.9%) |  |
| – Corning | 10 (0.4%) | 2 (0.4%) |  |
| – Good Samaritan | 349 (12.9%) | 46 (8.5%) |  |
| – Immanuel | 92 (3.4%) | 24 (4.4%) |  |
| – Lakeside | 901 (33.2%) | 121 (22.3%) |  |
| – Mercy | 103 (3.8%) | 26 (4.8%) |  |
| – Midlands | 193 (7.1%) | 31 (5.7%) |  |
| – MO Valley | 8 (0.3%) | 3 (0.6%) |  |
| – St. Elizabeth | 491 (18.1%) | 52 (9.6%) |  |
| – St. Francis | 286 (10.5%) | 136 (25.1%) |  |

**Bivariate Logistic regression**

| **Predictor** | **Comparison (vs Ref)** | **OR** | **95% CI (Lower, Upper)** | **p-value** |
| --- | --- | --- | --- | --- |
| **Age category** | 40–60 vs ≤40 | 0.36 | 0.25 – 0.52 | <0.001 |
|  | 60–80 vs ≤40 | 0.30 | 0.21 – 0.43 | <0.001 |
|  | 80+ vs ≤40 | 0.85 | 0.57 – 1.29 | 0.450 |
| **ER status** | Positive vs Negative | 0.46 | 0.37 – 0.58 | <0.001 |
| **Race** | Other vs Black | 0.62 | 0.33 – 1.16 | 0.138 |
|  | White vs Black | 0.43 | 0.29 – 0.65 | <0.001 |
| **Insurance** | Other vs Medicaid | 1.70 | 1.30 – 2.21 | <0.001 |
|  | Private vs Medicaid | 0.79 | 0.64 – 0.97 | 0.027 |
| **HER2 status** | Positive vs Negative | 1.48 | 1.15 – 1.88 | 0.002 |
| **PR status** | Positive vs Negative | 0.53 | 0.44 – 0.65 | <0.001 |

In bivariate logistic regression analyses, younger patients (≤40 years) had the highest risk of recurrence, with significantly lower odds observed in those aged 40–60 years (OR = 0.36, 95% CI: 0.25–0.52) and 60–80 years (OR = 0.30, 95% CI: 0.21–0.43), while patients aged ≥80 did not differ significantly from the youngest group. Tumor biology was strongly associated with recurrence: ER-positive (OR = 0.46, 95% CI: 0.37–0.58) and PR-positive (OR = 0.53, 95% CI: 0.44–0.65) tumors were less likely to recur, whereas HER2-positive tumors showed increased recurrence risk (OR = 1.48, 95% CI: 1.15–1.88). Racial differences were evident, with White patients experiencing significantly lower odds of recurrence compared to Black patients (OR = 0.43, 95% CI: 0.29–0.65), while the difference for patients categorized as Other was not statistically significant. Insurance status was also associated with recurrence, with privately insured patients having lower odds (OR = 0.79, 95% CI: 0.64–0.97) and those categorized as “Other” insurance types having higher odds (OR = 1.70, 95% CI: 1.30–2.21) relative to Medicaid.

Table 2: Multivariable mixed-effects logistic regression of recurrence

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor** | **Comparison (vs Ref)** | **OR** | **95% CI** | **p-value** |
| **Age category** | 40–60 vs ≤40 | 0.35 | 0.23 – 0.51 | <0.001 |
|  | 60–80 vs ≤40 | 0.31 | 0.21 – 0.46 | <0.001 |
|  | 80+ vs ≤40 | 0.89 | 0.57 – 1.38 | 0.594 |
| **ER status** | Positive vs Negative | 0.62 | 0.45 – 0.86 | 0.004 |
| **Race** | Other vs Black | 0.66 | 0.34 – 1.29 | 0.224 |
|  | White vs Black | 0.43 | 0.28 – 0.67 | <0.001 |
| **PR status** | Positive vs Negative | 0.72 | 0.55 – 0.95 | 0.020 |

In the multivariable mixed-effects logistic regression model, patient age, tumor receptor status, and race were significantly associated with recurrence. Compared with patients ≤40 years, those aged 40–60 years (OR = 0.35, 95% CI: 0.23–0.51) and 60–80 years (OR = 0.31, 95% CI: 0.21–0.46) had substantially lower odds of recurrence, while patients ≥80 years did not differ significantly (OR = 0.89, 95% CI: 0.57–1.38). ER-positive tumors were associated with 38% lower recurrence odds (OR = 0.62, 95% CI: 0.45–0.86), and PR positivity conferred a similar protective effect (OR = 0.72, 95% CI: 0.55–0.95). White patients experienced significantly reduced recurrence risk relative to Black patients (OR = 0.43, 95% CI: 0.28–0.67), whereas patients categorized as Other did not differ significantly (OR = 0.66, 95% CI: 0.34–1.29).

The model included a random intercept for facility to account for clustering of patients within treatment sites. The estimated facility-level variance was 0.194, corresponding to an intraclass correlation coefficient of 0.056, suggesting that approximately 6% of the variation in recurrence risk was attributable to between-facility differences after adjusting for patient characteristics. Model performance statistics indicated that fixed effects alone explained about 7.1% of the variance (marginal R²), while the combination of fixed and random effects explained 12.3% (conditional R²). This variance–components specification, with random intercepts only, implies that facility effects capture baseline differences in recurrence risk without altering the estimated associations of patient-level predictors.

A graph of a logistic model

AI-generated content may be incorrect.

The ROC curve demonstrates that the model has good discriminatory ability, with a marginal AUC of 0.77 (95% CI: 0.75–0.80) based on fixed effects alone and a conditional AUC of 0.83 (95% CI: 0.81–0.85) when facility-level random effects are included. These findings indicate that the selected patient- and tumor-level predictors meaningfully distinguish between patients with and without recurrence, while accounting for clustering by treatment facility further improves model discrimination.

**Discussion**

In this study, younger patients, particularly those aged 40 years and below, were more likely to experience recurrence, while patients in the 40–60 and 60–80-year categories had substantially lower odds compared with the youngest group. ER- and PR-positive tumors were both associated with significantly reduced recurrence risk, highlighting the protective effect of hormone receptor positivity in this cohort. Although HER2 status showed an association with recurrence in unadjusted analyses, it did not remain significant in the multivariable model once other predictors were included. Racial differences were evident, with White patients having significantly lower recurrence risk compared to Black patients, while patients classified as Other did not differ significantly from the reference group. The random intercept for facility suggested that about 6% of the variation in recurrence risk was explained by differences across treatment sites, indicating modest clustering effects. Despite relatively small R² values, the model demonstrated good discrimination, with an AUC of 0.77 for fixed effects and 0.83 when accounting for facility-level variation, showing that the selected predictors meaningfully distinguished patients with and without recurrence.