GOIANA JOURNAL OF MEDICINE

DOI 10.63162/V67N70E26720
ISSN 0034-9585
e26720
SCIENTIFIC ARTICLE - ORIGINAL

CLINICAL FACTORS ASSOCIATED WITH DEATH AND RECURRENCE IN INVASIVE BREAST CARCINOMA PATIENTS: PENALIZED LOGISTIC REGRESSION ANALYSIS IN A HOSPITAL COHORT

JUAREZ ANTÔNIO DE SOUSA[1], WALDEMAR NAVES DO AMARAL[1], ANDRÉ MAROCCOLO DE SOUSA[2], DÉBORA NASCIMENTO DIAS NEVES[3], LAURA DE SOUSA E SOUSA[4], LUIZA DE SOUSA E SOUSA[4], LETÍCIA CAIADO MADI[5]

1. PhD, Professor, Department of Gynecology and Obstetrics, Faculdade de Medicina, Universidade Federal de Goiás (UFG), Goiânia, GO, Brazil.
2. General Surgery Resident, Universidade de São Paulo (USP), São Paulo, SP, Brazil.
3. Physician, Universidade Federal de Goiás (UFG), Goiânia, GO, Brazil.
4. Medical Student, Universidade UniEvangélica, Anápolis, GO, Brazil.
5. Obstetrics and Gynecology Resident, Maternidade Aristina Cândida, Goiânia, GO, Brazil.

ABSTRACT

Introduction: Identifying clinical predictors of death and recurrence in breast cancer is essential for prognostic stratification and individualized treatment. In cohorts with few events, conventional statistical methods are inadequate, requiring penalized approaches such as Firth's regression. **Objective:** To identify clinical factors associated with death and tumor recurrence in patients with invasive breast carcinoma in a hospital cohort, using Firth's penalized logistic regression. **Methods:** Retrospective cohort study with 97 patients with confirmed invasive breast carcinoma, excluding ductal carcinoma in situ. Univariate analysis used Fisher's exact test. Firth's penalized logistic regression models were fitted for death and recurrence, expressed as odds ratios (OR) with 95% CI. Event concentration by subgroup was presented in a staging × T-component cross-table. **Results:** Six deaths (6.2%) and 9 recurrences (9.3%) were recorded. Advanced staging (III–IV) was strongly associated with death (OR = 57.67; 95%CI: 9.39–643.91; p < 0.001) and recurrence (OR = 73.29; 95%CI: 14.05–544.21; p < 0.001). T3–T4 component showed associations of similar magnitude. In a multivariable model for recurrence, staging III–IV maintained independent association (adjusted OR = 39.99; 95%CI: 7.41–290.74; p < 0.001) after adjustment for axillary approach and HER2. Patients with stage III–IV and T3–T4 tumors had death rates of 50% and recurrence rates of 70%. The adjusted model ROC curve reached AUC = 0.942. **Conclusion:** Advanced clinical staging and T3–T4 component are fundamental determinants of adverse prognosis in invasive breast carcinoma. Firth's penalized logistic regression proved appropriate for analyzing rare outcomes in small cohorts.

Keywords: Breast neoplasms, Neoplasm recurrence, Local, Prognosis, Logistic regression, Survival analysis.

INTRODUCTION

Breast cancer represents the most common malignant neoplasm among women in Brazil and worldwide, with a significant impact on female morbidity and mortality. The prognosis of the disease is multifactorial, being determined by clinical, histological, and molecular characteristics

of the tumor, in addition to the extent of the disease at diagnosis. Early identification of factors associated with death and recurrence is essential for risk stratification, individualization of therapeutic strategies, and appropriate clinical follow-up.[1]

From an epidemiological perspective, clinical stage at diagnosis is recognized as the main predictor of overall survival in breast cancer. Patients diagnosed at early stages (I–II) have five-year survival rates close to 90%, while those with locally advanced (III) or metastatic (IV) disease have substantially worse prognoses. In addition to staging, factors such as tumor size, axillary lymph node involvement, histological grade, and molecular profile — especially the triple-negative subtype — are independent predictors of recurrence and mortality widely documented in the literature.[2]

Conducting robust inferential analyses in small hospital cohorts poses relevant methodological challenges. When the number of observed events is low, estimates obtained by conventional logistic regression are susceptible to complete or quasi-complete separation bias, with consequent instability of the coefficients and excessively wide confidence intervals. In these situations, Firth penalized logistic regression is recommended because of its ability to produce unbiased estimates even in the presence of data separation, being particularly suitable for studies with few events per covariate.[3]

The present study aimed to identify clinical factors associated with death and tumor recurrence in patients with invasive breast carcinoma followed in a hospital cohort, using statistical methods appropriate for samples with a low number of events.[4]

METHODS

Study Design and Population

A retrospective observational cohort study was conducted at a private referral hospital for oncology in the state of Goiás, in the Central-West region of Brazil. Patients with a histologically confirmed diagnosis of invasive breast carcinoma were included. The following were excluded: (1) cases of ductal carcinoma in situ (DCIS, stage 0), as they do not represent invasive disease; (2) patients with missing essential clinical data. The final sample consisted of 97 patients.

Variables and Outcomes

The primary outcomes were death from any cause and tumor recurrence (local, regional, or distant), both treated as binary variables. The explanatory variables included: clinical stage (I–II vs. III–IV), T component of the TNM classification (T1–T2 vs. T3–T4), histological type (IDC, ILC, inflammatory carcinoma, and others), tumor grade (G1 vs. G2–G3), axillary status (negative vs. positive), molecular profile (HR-positive, HER2-positive, triple-negative subtype), and treatment modalities performed (chemotherapy, hormone therapy, and radiotherapy).

Statistical Analysis

Descriptive analysis was performed using absolute and relative frequencies for categorical variables. Univariate analysis used Fisher's exact test to assess the association between clinical variables and the outcomes, due to the small number of events. Variables with a statistically significant association or potential clinical relevance were included in the regression models.

Firth penalized logistic regression models were fitted for both outcomes. This method reduces estimation bias in situations of complete or quasi-complete separation of the data, a common condition in datasets with few events. Results were expressed as odds ratios (ORs) with 95% confidence intervals

(95% CIs). In an exploratory multivariable model for recurrence, stage, axillary status, and HER2 status were included as covariates. The discriminatory ability of the models was assessed by the area under the ROC curve (AUC). All analyses were performed using R software (version 4.x; R Foundation for Statistical Computing, Vienna, Austria), with a significance level of 5%.

RESULTS

Sample Characteristics

A total of 97 patients with invasive breast carcinoma were included, after the exclusion of 2 cases of DCIS. During follow-up, 6 deaths (6.2%) and 9 recurrences (9.3%) were recorded. The clinical characteristics of the sample are presented in Table 1.

Table 1 : Clinicopathological characteristics of the sample (n = 97).

Variable	n	%
Clinical Stage		
I–II	87	89.7
III–IV	10	10.3
T component (TNM)		
T1–T2	86	88.7
T3–T4	11	11.3
Histological type		
Invasive ductal carcinoma (IDC)	80	82.5
Invasive lobular carcinoma (ILC)	7	7.2
Inflammatory carcinoma	2	2.1
Other subtypes	8	8.2
Tumor grade		
G1	14	14.4
G2	69	71.1
G3	14	14.4
Axillary approach		
Negative	69	71.1
Positive	28	28.9
Molecular profile		
HR-positive (ER or PR)	75	77.3
HER2-positive	19	19.6
Triple-negative subtype	15	15.5
Treatments performed		
Chemotherapy	56	57.7
Neoadjuvant	32	33.0
Adjuvant	22	22.7
Hormone therapy	74	76.3
Radiotherapy	75	77.3
Outcomes		
Death	6	6.2
Recurrence	9	9.3

IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; HR: hormone receptors.

Univariate Analysis — Fisher's Exact Test

In the univariate analysis for the outcome of death (Table 2), a statistically significant association was observed with variables related to tumor extent: TNM risk, N component, clinical stage, T component, and hormone therapy (all p < 0.001), as well as chemotherapy (p = 0.036) and tumor grade (p = 0.043). Axillary status (p = 0.052) and histological type (p = 0.083) showed borderline associations.

Table 2 : Fisher's exact test — association between clinical variables and death (n = 97).

Variable	N	p-value
TNM risk	97	< 0.001
TNM N component	97	< 0.001
Clinical stage	97	< 0.001
TNM T component	97	< 0.001
Hormone therapy	97	< 0.001
Chemotherapy	97	0.036
Tumor grade	97	0.043
Axillary approach	97	0.052
Histological type	97	0.083
Reconstruction	97	0.481
Radiotherapy	97	0.621
Educational level	97	0.679
Family history	97	1.000
HER2	97	1.000

For the outcome of recurrence (Table 3), significant associations included the N component, TNM risk, clinical stage, T component, hormone therapy, chemotherapy, and axillary status. HER2 showed a borderline association (p = 0.066).

Table 3: Fisher's exact test — association between clinical variables and recurrence (n = 97).

Variable	N	p-value
TNM N component	97	< 0.001
TNM risk	97	< 0.001
Clinical stage	97	< 0.001
TNM T component	97	< 0.001
Hormone therapy	97	< 0.001
Chemotherapy	97	0.009
Axillary approach	97	0.014
HER2	97	0.066
Tumor grade	97	0.135
Educational level	97	0.161
Histological type	97	0.368
Radiotherapy	97	0.430
Reconstruction	97	1.000
Family history	97	1.000

Firth Penalized Logistic Regression — Univariate Models

The univariate Firth models (Table 4) showed a strong association of stage III–IV with death (OR = 57.67; 95% CI: 9.39–643.91; p < 0.001) and recurrence (OR = 73.29; 95% CI: 14.05–544.21; p < 0.001).

The T3–T4 component showed associations of similar magnitude for both outcomes (death: OR = 49.36; 95% CI: 8.24–539.63; p < 0.001; recurrence: OR = 57.67; 95% CI: 11.70–395.81; p < 0.001). Regarding histological type, the "Other" group (vs. IDC) was significantly associated with death (OR = 8.16; 95% CI: 1.53–44.22; p = 0.016), whereas ILC showed no significant difference for either outcome. The wide confidence intervals reflect the small number of events.

Table 4: Firth Penalized Logistic Regression — Univariate Models for Death and Recurrence.

Model	Comparison	OR (95% CI)	p-value
Outcome: Death			
Clinical stage	III–IV vs I–II	57.67 (9.39–643.91)	< 0.001
T component	T3–T4 vs T1–T2	49.36 (8.24–539.63)	< 0.001
Histological type	ILC vs IDC	1.48 (0.01–17.56)	0.811
Histological type	Other vs IDC	8.16 (1.53–44.22)	0.016
Outcome: Recurrence			
Clinical stage	III–IV vs I–II	73.29 (14.05–544.21)	< 0.001
T component	T3–T4 vs T1–T2	57.67 (11.70–395.81)	< 0.001
Histological type	ILC vs IDC	0.76 (0.01–7.66)	0.854
Histological type	Other vs IDC	4.22 (0.88–17.90)	0.069

IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; OR: odds ratio; 95% CI: 95% confidence interval.

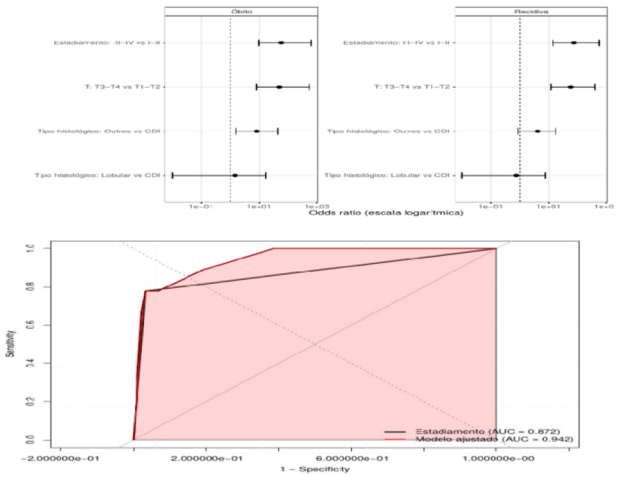

Figure 1: Forest plot of Firth penalized logistic regression models for the outcomes of death and recurrence (upper panel), and ROC curves comparing the model with clinical stage alone (AUC = 0.872) versus the model adjusted for clinical stage, axillary approach, and HER2 (AUC = 0.942; lower panel). The dashed line indicates OR = 1.

Exploratory Multivariable Model for Recurrence

In the multivariable model for recurrence (Table 5), stage III–IV remained strongly associated with

the outcome after adjustment for axillary approach and HER2 status (adjusted OR = 39.99; 95% CI: 7.41–290.74; p < 0.001). Positive axillary approach (OR = 2.01; 95% CI: 0.27–12.74; p = 0.467) and HER2-positive status (OR = 2.62; 95% CI: 0.35–19.21; p = 0.335) did not show statistically significant independent associations after adjustment.

Table 5: Exploratory Multivariable Model (Firth Regression) — Predictors of Recurrence.

Variable	Adjusted OR (95% CI)	p-value
Clinical stage III–IV vs. I–II	39.99 (7.41–290.74)	< 0.001
Positive axillary approach	2.01 (0.27–12.74)	0.467
HER2-positive status	2.62 (0;35–19.21)	0.335

OR: adjusted odds ratio; 95% CI: 95% confidence interval.

Concentration of Events by Subgroup

Table 6 and Figure 2 present the distribution of death and recurrence according to the main clinical subgroups. Patients with stage III–IV disease and T3–T4 tumors had death and recurrence proportions of 50% (5/10) and 70% (7/10), respectively, compared with less than 2% for both outcomes among patients with early-stage disease (stage I–II, T1–T2). The concentration of events in the triple-negative subgroup was also substantial — 26.7% for deaths and 33.3% for recurrences, versus 2.4% and 4.9%, respectively, in the other subgroups.

Table 6: Proportion of Death and Recurrence According to Clinical Subgroups (n = 97).

Subgroup	n	Deaths n (%)	Recurrences n (%))	p-value*
Stage I–II	87	1 (1.1)	2 (2.3)	< 0.001
Stage III–IV	10	5 (50.0)	7 (70.0)	< 0.001
T1–T2	86	1 (1.2)	2 (2.3)	< 0.001
T3–T4	11	5 (45.5)	6 (54.5)	< 0.001
Negative axilla	69	1 (1.4)	2 (2.9)	0.052 / 0.014
Positive axilla	28	5 (17.9)	7 (25.0)	0.052 / 0.014
Triple-negative — No	82	2 (2.4)	4 (4.9)	< 0.001
Triple-negative — Yes	15	4 (26.7)	5 (33.3)	< 0.001

*p-value by Fisher's exact test; values separated by a slash correspond to the p-value for death / p-value for recurrence, respectively.

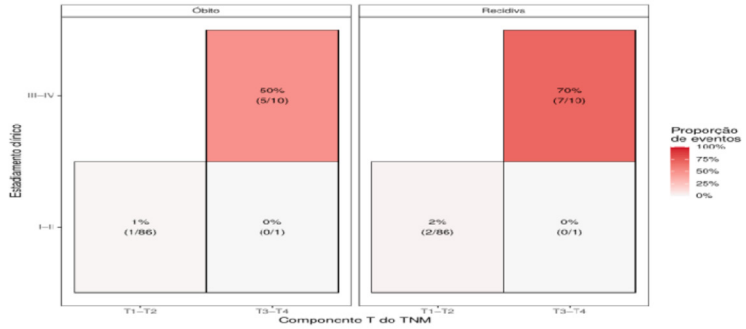

Figure 2: Proportion of death and recurrence according to clinical stage and the T component of the TNM system. The values in each cell represent the proportion of events in the corresponding subgroup. A clear concentration of events was observed among patients with stage III–IV disease and T3–T4 tumors.

DISCUSSION

The results of this study confirm the central role of clinical stage as the main prognostic determinant in invasive breast carcinoma, with associations of large magnitude for both death and recurrence, which were maintained after multivariable adjustment. Patients with stage III–IV disease had approximately 58- and 73-fold higher odds of death and recurrence, respectively, compared with patients with stage I–II disease — a finding consistent with the established oncology literature and biologically plausible, given the greater tumor burden and therapeutic limitations inherent to advanced disease.[5]

The T component of the TNM system showed associations of similar magnitude to overall stage, which is expected given the strong correlation between these two parameters. T3–T4 tumors, due to their greater tumor volume and higher probability of lymph node involvement and distant spread, are associated with poorer locoregional control and a higher risk of systemic recurrence. Cross-analysis of clinical stage × T component revealed that adverse events were specifically concentrated in the subgroup with both unfavorable factors: proportions of 50% for death and 70% for recurrence among patients with stage III–IV disease and T3–T4 tumors.[6]

The "Other" histological subtype showed a significant association with death (OR = 8.16; p = 0.016), a result that warrants careful interpretation. This group is heterogeneous and includes Paget and medullary carcinomas — subtypes with distinct prognoses. The wide confidence intervals observed in the estimates for histological type reflect both the small number of events and the within-group biological heterogeneity, limiting definitive conclusions.[7]

In the exploratory multivariable model for recurrence, stage III–IV remained independently associated with the outcome after adjustment for axillary approach and HER2 status. The lack of statistical significance for positive axillary approach and HER2 in the adjusted model does not imply a lack of clinical relevance of these factors, but rather reflects the limited statistical power of this sample. With only 9 recurrence events, the model's ability to detect independent effects of additional covariates is substantially reduced.[8]

The adjusted model achieved an AUC of 0.942 for predicting recurrence, suggesting excellent discriminatory ability. However, this value should be interpreted with caution: in small samples with

a low number of events, the AUC tends to be overestimated, and external validation of the model would be necessary before any clinical application. The improvement in AUC from 0.872 (model with clinical stage alone) to 0.942 (adjusted model) indicates that the inclusion of additional covariates adds discriminatory value, although again subject to the limitations of the sample context.[9]

From a methodological perspective, the use of Firth penalized logistic regression was appropriate for the context of this study. The quasi-complete separation of the data — resulting from the strong association between advanced stage and the occurrence of events in a sample with few cases — would render conventional logistic regression estimates unstable or even incalculable. The Firth method addresses this problem by introducing a penalty into the likelihood function, producing estimates with reduced bias even in situations of separation.

The main limitations of the study are the small sample size and the low number of events, which impose restrictions on statistical power and the stability of multivariable estimates. The retrospective nature of the study and its single-center setting limit the generalizability of the findings. Prospective studies with larger series and prolonged follow-up are needed to confirm the identified associations and to develop prognostic models with external validation.[10]

CONCLUSION

O estadiamento clínico avançado (III–IV) e o componente tumoral T3–T4 constituíram os principais determinantes do prognóstico adverso nesta coorte de pacientes com carcinoma mamário invasivo, associando-se independentemente a maiores riscos de óbito e recidiva tumoral. A regressão logística penalizada de Firth mostrou-se abordagem estatisticamente adequada para o estudo de desfechos raros em séries clínicas de pequeno porte, permitindo estimativas estáveis na presença de separação quase completa dos dados.

Os achados reforçam a relevância do diagnóstico precoce e do estadiamento preciso como ferramentas fundamentais para a estratificação prognóstica e orientação das decisões terapêuticas em oncologia mamária. A ampliação da casuística e o seguimento prolongado de séries institucionais são necessários para a consolidação de modelos prognósticos validados e aplicáveis à realidade clínica regional.

REFERENCES

1. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Estimativa 2023: Incidência de Câncer no Brasil. Rio de Janeiro: INCA; 2022.

2. 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.

3. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nat Rev Dis Primers. 2019;5(1):66.

4. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27-38.

5. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21(16):2409-19.

6. Heinze G, Ploner M. Fixing the nonconvergence bug in logistic regression with SPLUS and SAS. Comput Methods Programs Biomed. 2003;71(2):181-7.

7. Rakha EA, El-Sayed ME, Lee AH, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. J Clin Oncol. 2008;26(19):3153-8.

8. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of

targeted therapies. J Clin Invest. 2011;121(7):2750-67.

9. Gradishar WJ, Moran MS, Abraham J, Aft R, Agnese D, Allison KH, et al. NCCN Guidelines® Insights: Breast Cancer, Version 4.2023. J Natl Compr Canc Netw. 2023;21(6):594-608.

10. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2023. Disponível em: https://www.R-project.org/.

MAILING ADDRESS

JUAREZ ANTÔNIO DE SOUSA
Rua 13, 380 ap 5, Setor Oeste, Goiânia/GO, Brazil.
E-mail: juarez_antonio@ufg.br

EDITORIAL AND REVIEW

Authors:
Juarez Antônio de Sousa - http://lattes.cnpq.br/4484429936026476 - https://orcid.org/0000-0001-5986-7926

Waldemar Naves do Amaral - http://lattes.cnpq.br/4092560599116579 - https://orcid.org/0000-0002-0824-1138

André Maroccolo de Sousa - http://lattes.cnpq.br/4379236388126901 - https://orcid.org/0009-0004-5613-1410

Débora Nascimento Dias Neves - http://lattes.cnpq.br/5830809552139767 - https://orcid.org/0000-0003-1280-5415

Laura de Sousa e Sousa - http://lattes.cnpq.br/5775215940326611- https://orcid.org/0009-0004-9019-8846

Luiza de Sousa e Sousa - http://lattes.cnpq.br/9882323304990808 - https://orcid.org/0009-0007-3855-8682

Letícia Caiado Madi - http://lattes.cnpq.br/3769545831010905 - https://orcid.org/0000-0002-2038-1379

Library Review: Izabella Goulart
Spell Check: Dario Alvares
Translation: Soledad Montalbetti
Received: 31/05/26. Accepted: 14/07/26. Published in: 17/08/2026.