Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.36641

Survival Analysis of Male Breast Cancer Patients Based on SEER Data

Yanhai Xiao

Department of Mathematics and Statistics, Brock University, St. Catharines, Canada
wa24iq@brocku.ca

Abstract. Male breast cancer is uncommon, but its clinical management remains challenging because evidence is still far less abundant than that for female breast cancer. This study used data from the SEER Program to evaluate prognostic factors in male breast cancer and to compare the performance of Cox proportional hazards and accelerated failure time (AFT) models. After data cleaning and variable construction, 1,743 male patients diagnosed between 2000 and 2019 were included in the final analytic cohort. The dataset contained demographic information, tumor stage, receptor status, treatment variables, survival time, and outcome status. According to the above results, the 1-year, 3-year and 5-year survival rates are 93.7%, 82.7% and 71.5%, respectively. The test for proportional hazards showed that some variables did not meet the requirements of the Cox model, so an AFT model was selected. We can find that survival analysis of SEER data can provide statistical support for the prognosis prediction of patients.

Keywords: Male Breast Cancer, SEER, Survival Analysis

1. Introduction

Male breast cancer (MBC) accounts for only a small proportion of all breast cancer cases, but it remains an important clinical problem because affected patients are often diagnosed at older ages and with more advanced disease than female patients [1-6]. In routine practice, many therapeutic recommendations for MBC are still extrapolated from the evidence base established for female breast cancer, reflecting the limited number of male-specific prospective studies [6-9]. As a result, there is continuing interest in clarifying the prognostic profile of MBC using large population-based datasets.

Another challenge in MBC research is the difficulty of assembling sufficiently large cohorts from single institutions. Because of its rarity, hospital-based series frequently have limited sample sizes, which can reduce statistical power and restrict the evaluation of multiple prognostic variables. Population registries such as the Surveillance, Epidemiology, and End Results (SEER) Program provide a useful alternative because they contain a broad range of demographic, tumor, treatment, and outcome information from a large and diverse population [2, 7]. These data make it possible to examine survival patterns more systematically and to explore how clinical characteristics and treatment strategies are associated with prognosis.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.36641

Most survival studies in oncology rely heavily on the Cox proportional hazards model. Although this model is interpretable, its validity depends on the proportional hazard's assumption. When the effect of a covariate changes over time, the Cox framework may provide an incomplete description of the survival process [4, 5]. In such settings, accelerated failure time (AFT) models can be valuable because they describe how covariates lengthen or shorten survival time directly and do not require proportional hazards in the same way [10].

The present study used SEER data to examine the survival of male breast cancer patients diagnosed between 2000 and 2019. The objectives were twofold. First, this study aimed to identify demographic, pathological, and treatment-related factors associated with prognosis. Second, it aimed to compare the performance of Cox proportional hazards and AFT models in order to assess whether a parametric time-based approach provides advantages for this dataset. By doing so, the study seeks to contribute additional evidence for risk stratification and methodological decision-making in MBC research.

2. Methods

2.1. Data source and cleaning

The data come from the SEER database of the National Cancer Institute; this is a cancer registry system based on a population that collects cancer incidence, prevalence and survival data for about 34.6 per cent of the U. S. Population. Data from 2000 to 2019 on male breast cancer patients have been selected. The first set of data were the patient, tumour, treatment and survival records.

The three periods of the data cleaning are as follows: First, variable selection and recording were performed based on the research goals, and some categorical variables were recoded. Next, for the missing values in the key variables, I used a strategy of removing records with missing values to ensure the accuracy of the analysis. The final Data Set has 1743 entries. The last one is called the survival-time process. Survival time was in months, and the state of being alive or dead was 0 or 1, respectively. To avoid a zero-survival time, all the records with a survival time were rounded up to one month.

2.2. Statistical models

Kaplan-Meier Survival Curves and Log-rank Test: Kaplan-Meier (KM) survival curves are employed to estimate and plot the probability of survival over time for different patient groups [3]. A log-rank test is performed to determine if there is a statistically significant difference in survival curves among multiple groups, and thus, it can be concluded that different categories of a variable effect survival.

Cox Proportional Hazards (Cox PH) Model: The Cox PH model is a semi-parametric regression model which assesses the effect of multiple covariates on survival time [4]. It requires that the Hazard Ratio (HR) remains constant over time for each covariate. The HR value output by the model shows the relative change in the risk of an event when a covariate increases by one unit after controlling other variables. The cox.zph function is used to test the hazards assumption.

Accelerated Failure Time (AFT) Model: AFT models are a commonly used class of survival analysis models which model the logarithm of survival time and assume that covariates have a multiplicative effect on survival time [5]. AFT models focus on Time Ratios (TR). This study used two common AFT models, such as Weibull AFT Model and Log-normal AFT Model.

DOI: 10.54254/2753-8818/2026.36641

Model Comparison: We evaluated the goodness-of-fit and predictive ability of the models through the Log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and C-index.

3. Results

3.1. Data overview

There are 1,743 male breast cancer (MBC) patients and 535 deaths among them in the dataset. The Median Follow-up Time was 46 months. A relatively large sample size is available for further survival analysis, and thus the accuracy of statistical estimation will improve. The 1-year, 3-year and 5-year survival rates for MBC patients in the overall survival analysis were 93.7%, 82.7% and 71.5%. The survival curve showed a continuous decrease in the probability of survival with an increase in the follow-up period; therefore, the long-term survival of MBC patients remains a problem (Table 1).

Table 1. Kaplan–meier survival curves stratified by T stage and metastatic status in male breast cancer patients

Time (Months)	Survival Rate	95% CI Lower	95% CI Upper
12	0.937	0.925	0.949
36	0.827	0.807	0.846
60	0.715	0.690	0.741

The results of the Log-rank test show that both T stage grouping (Chi-sq = 76.2, df = 2, $p < 2e\text{-}16$) and metastatic status (Chi-sq = 178, df = 1, $p < 0.001$) were associated with differences in patient survival.

3.2. Cox proportional hazards model

The Cox Proportional Hazards model analysis results showed that several factors were significantly associated with the survival risk of MBC patients. Compared to the <60 years group, the risk of death increased in the 60-79 years group and 80+ years group, which had HRs of 1.87 and 6.23, and compared to the T0-T1 group, the risk of death significantly increased in the T2 group and T3-T4 group, with HRs of 1.82 and 2.26. The risk of death in the N2-N3 group was also higher than in the N0-N1 group, with an HR of 1.52, and the risk of death in the M1 group was higher than in the M0 group, with an HR of 1.82. Otherwise, surgery, radiation, and chemotherapy all significantly reduced the risk of death (Table 2).

Table 2. Cox proportional hazards model results

Variable	Coefficient	Hazard Ratio (HR)	95% CI Lower	95% CI Upper	p-value
Surgery: Yes	-1.08	0.34	0.26	0.44	< 0.001
Age: 80+ years	1.83	6.23	3.89	10.01	< 0.001
T stage: T2	0.60	1.82	1.50	2.22	< 0.001
Chemotherapy: Yes	-0.65	0.52	0.42	0.65	< 0.001

DOI: 10.54254/2753-8818/2026.36641

Table 2. (continued)

T stage: T3-T4	0.81	2.26	1.69	3.01	< 0.001
Age: 60-79 years	0.62	1.87	1.46	2.38	< 0.001
M stage: M1	0.60	1.82	1.32	2.50	< 0.001
N stage: N2-N3	0.42	1.52	1.12	2.07	0.007
Radiation: Yes	-0.25	0.78	0.62	0.96	0.022
HER2: Positive	0.22	1.24	0.93	1.66	0.147
PR: Positive	-0.24	0.79	0.56	1.11	0.174
N stage: N1	0.15	1.16	0.94	1.43	0.174
ER: Positive	-0.29	0.75	0.41	1.34	0.326
Race: Other	0.13	1.14	0.82	1.59	0.448
Race: Black	-0.08	0.93	0.71	1.21	0.583

The proportional hazards assumption test for the Cox PH model showed that variables ER, PR and surgery violated the proportional hazards assumption; thus, their hazard ratios may vary over time, and AFT models may be more appropriate for describing their impact on survival time.

3.3. Accelerated failure time models

We could use Weibull and Log-normal AFT models, because the proportional hazards assumption was violated for some variables in the Cox PH model. The results of AFT models are shown as Time Ratios (TR). TR > 1 means longer survival time, and TR < 1 means shorter survival time.

The Weibull AFT model results showed consistent trends with the Cox PH model. Compared to the <60 years group, the survival time was shorter in the 60-79 years group and 80+ years group, with TRs of 0.60 and 0.23, and compared to the T0-T1 group, the survival time was significantly shorter in the T2 group and T3-T4 group, which had TRs of 0.62 and 0.52. The survival time in the M1 group was also shorter, with a TR of 0.62. On the other hand, surgery, radiation, and chemotherapy all let survival time longer (Table 3).

Table 3. Weibull AFT model results

Variable	Coefficient	Time Ratio (TR)	95% CI Lower	95% CI Upper	p-value
Surgery: Yes	0.86	2.37	1.91	2.93	< 0.001
Age: 80+ years	-1.46	0.23	0.16	0.34	< 0.001
T stage: T2	-0.48	0.62	0.53	0.72	< 0.001
Chemotherapy: Yes	0.52	1.67	1.40	2.01	< 0.001
T stage: T3-T4	-0.65	0.52	0.41	0.66	< 0.001
Age: 60-79 years	-0.51	0.60	0.49	0.74	< 0.001
M stage: M1	-0.48	0.62	0.48	0.80	< 0.001
N stage: N2-N3	-0.33	0.72	0.56	0.92	0.010
Radiation: Yes	0.21	1.23	1.03	1.47	0.021

DOI: 10.54254/2753-8818/2026.36641

Table 3. (continued)

HER2: Positive	-0.17	0.84	0.66	1.06	0.147
N stage: N1	-0.12	0.89	0.75	1.06	0.187
PR: Positive	0.18	1.20	0.91	1.59	0.195
ER: Positive	0.27	1.30	0.81	2.10	0.275
Race: Other	-0.10	0.90	0.69	1.18	0.451
Race: Black	0.06	1.06	0.86	1.32	0.572

The Log-normal AFT model results also showed similar significant factors, such as the Cox PH model. Compared to the <60 years group, the survival time was shorter in the 60-79 years group and 80+ years group, with TRs of 0.57 and 0.23, and compared to the T0-T1 group, the survival time was shorter in the T2 group and T3-T4 group, with TRs of 0.58 and 0.48. The survival time in the M1 group was also shorter, which had a TR of 0.51. Surgery, radiation, and chemotherapy all made survival longer (Table 4).

Table 4. Log-normal AFT model results

Variable	Coefficient	Time Ratio (TR)	95% CI Lower	95% CI Upper	p-value
Surgery: Yes	1.17	3.23	2.46	4.23	< 0.001
Age: 80+ years	-1.47	0.23	0.14	0.37	< 0.001
Chemotherapy: Yes	0.62	1.85	1.51	2.28	< 0.001
T stage: T2	-0.54	0.58	0.49	0.70	< 0.001
T stage: T3-T4	-0.74	0.48	0.36	0.63	< 0.001
Age: 60-79 years	-0.56	0.57	0.46	0.71	< 0.001
M stage: M1	-0.67	0.51	0.37	0.70	< 0.001
Radiation: Yes	0.29	1.33	1.09	1.62	0.005
N stage: N2-N3	-0.39	0.68	0.50	0.92	0.012
ER: Positive	0.63	1.87	1.08	3.23	0.025
PR: Positive	0.21	1.24	0.89	1.72	0.201
HER2: Positive	-0.13	0.88	0.67	1.15	0.344
Race: Black	0.11	1.11	0.86	1.44	0.415
N stage: N1	-0.05	0.95	0.78	1.16	0.599
Race: Other	-0.06	0.94	0.68	1.29	0.699

3.4. Model comparison

We have compared the Cox PH model, Weibull AFT model and Log-normal AFT model, and the results are shown in the table below. Based on the AIC and BIC values, both Weibull AFT and Log-normal AFT models outperformed the Cox PH model; therefore, among the parametric models, they showed better fit to the data. The concordance index of the Log-normal AFT model was slightly

DOI: 10.54254/2753-8818/2026.36641

higher than those of the other two models, and thus had a better predictive capability for patient survival (Table 5).

Table 5. Model performance comparison

Model	Log-likelihood (LogLik)	AIC	BIC	Concordance Index (C-index)
Cox Proportional Hazards	-3410.06	6850.13	6932.08	0.7477
Weibull AFT	-3125.35	6284.71	6377.59	0.7472
Log-normal AFT	-3136.80	6307.61	6400.48	0.7500

4. Discussion

We performed a survival analysis of MBC patients from the SEER database in this paper, and age, tumour stage and treatment mode were associated with poor outcomes in MBC. Age is an independent risk factor; that is to say, older people have a much higher death rate than younger people.

Treatment modes of surgery, radiation therapy and chemotherapy all showed the same good survival rates. Surgery had the best results, and radiation and chemotherapy were still effective.

Based on the selection of a model, the proportional hazards assumption test for the Cox PH model indicated that ER, PR and surgery variables did not meet this assumption. According to the AIC and BIC values, both Weibull AFT and Log-normal AFT models are better than the Cox PH model. The Log-normal AFT model had a marginally higher concordance index.

We found that age is the main risk factor, and among surgery, radiotherapy and chemotherapy, these have shown better survival rates for patients. Older age and advanced stage are associated with a poorer prognosis of male breast cancer; so, it frequently spreads to lymph nodes or other organs.

This study still has several limitations. As a registry-based analysis, it is subject to residual confounding and the limitations of routinely collected data, and complete-case analysis necessarily excluded records with missing values, which may have introduced some selection bias. SEER also lacks certain potentially relevant clinical details, such as comorbidities, recurrence information, endocrine therapy details, and more granular molecular data. Finally, although the AFT models performed well, their adequacy still depends on distributional assumptions, and model choice should therefore be interpreted in the context of the research objective.

Despite these limitations, the study contributes useful evidence in two ways. Substantively, it identifies age, stage, and treatment-related variables as major correlates of prognosis in male breast cancer. Methodologically, it shows that time-based parametric survival models can perform competitively, and in some respects better than the conventional Cox approach, in SEER-derived MBC data.

5. Conclusion

Using SEER data from 2000 to 2019, this study evaluated prognostic factors and compared survival-modeling strategies in male breast cancer. Age, grouped T stage, grouped N stage, grouped M stage,

DOI: 10.54254/2753-8818/2026.36641

surgery, radiation, and chemotherapy were all important correlates of survival. Older age and more advanced disease were associated with poorer outcomes, whereas treatment variables were associated with longer survival and lower mortality risk.

The comparison between modeling approaches showed that the Cox model remained clinically interpretable, but its assumptions were not fully satisfied for all covariates. Both Weibull and Log-normal AFT models demonstrated better information-criterion performance, and the Log-normal AFT model achieved the strongest discrimination. These results suggest that when analyzing male breast cancer survival in registry data, AFT models deserve consideration alongside the Cox framework, especially when the proportional hazards assumption is questionable.

References

[1] American Cancer Society. (n.d.). Male breast cancer. *American Cancer Society*. https://www.cancer.org/cancer/types/breast-cancer-in-men.html

[2] National Cancer Institute. (n.d.). Surveillance, epidemiology, and end results (SEER) program: Survival statistics. *National Cancer Institute*. https://seer.cancer.gov/statistics/types/survival.html

[3] Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*, 457–481.

[4] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B, 34*, 187–220.

[5] Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: Multivariate data analysis—An introduction to concepts and methods. *British Journal of Cancer, 89*, 431–436.

[6] Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine, 11*, 1879–1891.

[7] Liu, N., Johnson, K. J., & Ma, C. X. (2018). Male breast cancer: An updated surveillance, epidemiology, and end results data analysis. *Clinical Breast Cancer, 18*, e997–e1002.

[8] Xie, J., Ying, Y. Y., Xu, B., Li, Y., Zhang, X., Li, C., He, J., Wang, X., & Yang, H. (2019). Metastasis pattern and prognosis of male breast cancer patients in the US: A population-based study from the SEER database. *Therapeutic Advances in Medical Oncology, 11*, 1758835919889003.

[9] Yao, N., Shi, W., Liu, T., Siyin, S. T., Wang, W., Duan, N., Xu, G., & Yeung, S. J. (2022). Clinicopathologic characteristics and prognosis for male breast cancer compared to female breast cancer. *Scientific Reports, 12*, 69.

[10] Zheng, G., & Leone, J. P. (2022). Male breast cancer: An updated review of epidemiology, clinicopathology, and treatment. *Journal of Oncology, 2022*, 1734049.