STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **12**, September 2024, pp 1492–1512.
Published online in International Academic Press (www.IAPress.org)

Breast cancer survival analysis and machine learning to predict the impact of different treatments

Amal Elnawasany*, BenBella S. Tawfik, Mohamed Makhlouf

Information Systems Department,Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt

Abstract Breast cancer is the most common form of cancer among women, impacting approximately one million women worldwide. New treatments are being developed yearly, improving breast cancer patients' survival rates. To explore the impact of different treatments, we conducted this study using data from the Surveillance, Epidemiology, and End Results (SEER) database. The study employed Kaplan-Meier analysis to examine breast cancer-specific survival (BCSS) and overall survival (OS) rates across various treatment options, including 'chemotherapy', 'radiotherapy, 'both therapies', and 'no therapy'. The log-rank test was also utilized to assess the statistical significance of differences observed between multiple survival curves. We found that recommended treatment for most breast cancer cases, based on BCSS analysis, is the combination of 'both' chemotherapy and radiotherapy. On the other hand, according to OS analysis, 'radiotherapy only' or 'in conjunction with chemotherapy' is the superior treatment for most breast cancer cases. They are often preferred over 'chemotherapy only' for most breast cancer patients. Machine learning was used to develop ten models predicting the survivability for OS and BCSS. C5.0 algorithm consistently achieves strong overall performance. It achieves high accuracy 0.98 and sensitivity of 0.99 for both OS and BCSS, reasonably RMSE of (0.14, 0.15 for BCSS and OS, respectively), and good ROC score of (0.91, 0.88 for BCSS and OS, respectively).

Keywords Breast cancer, Survival analysis, Chemotherapy, Radiotherapy, overall survival (OS), breast cancer-specific survival (BCSS)

DOI: 10.19139/soic-2310-5070-2027

1. Introduction

Breast cancer is a widely prevalent disease, accounting for approximately one out of every eight cancer diagnoses globally. In 2020, million new breast cancer cases were recorded, resulting in 685,000 deaths worldwide. By 2040, breast cancer cases are expected to increase exceeding 3 million new cases annually (a 50% increase) and over 1 million deaths per year (about 32% increase) [1] . Among women, breast, lung, and colorectal cancers are the three most common types, collectively accounting for half of all newly diagnosed cancer cases [2] . Breast cancer only represents 25% of all new cancer diagnoses in women [3] . Promisingly, there has been a significant decrease of 43% in breast cancer death rates among women from 1989 to 2020, primarily attributed to improved early-stage detection and diagnosis [4] . Early cancer detection plays a crucial role in increasing the chanced of patient survival. Despite advancements in breast cancer awareness and treatment, mortality rates still remain relatively high [5] . Cancer can aggressive and pose a life-threatening risk, but the chances of survival are considerably higher with early detection and timely treatment [6, 7] A study [8] compared the efficiency of breast-conserving therapy and mastectomy on OS using SEER database. It showed that BCT patients have improved survival compared with mastectomy in early-stage breast cancer patients.it encourage patients to receive BTC rather than mastectomy in early stage.

*Correspondence to: Amal Elnawasany (Email: aml.elnawasany@ci.suez.edu.eg). Information Systems Department, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

A retrospective cohort study [9] for patients with stage II/III BC, aged ¿= 55 compared patients aged ¿= 75 with younger patients. The study found that BCSS was approximately 95% in all ages but it was better in young patients with triple negative and HER+. BCSS was. The OS was poorer in the older group for all subtypes. Chemotherapy improved OS in different subtypes.

A study [10] investigated risk factors treatment impacts for male breast cancer. It suggested that Surgery and chemotherapy are recommended for early stages (I–III), however omitting other treatments, radiation or chemotherapy, worsened outcomes. For stage IV, both chemotherapy and radiation is suggested and improved BCSS.

A study [11] determined the comparative effect of chemotherapy and non-chemotherapy on breast cancer patients with different stages. The study aimed to clarify the potential differences in treatment effectiveness depending on the cancer stage. In the model, the relationship between the patient's marital status, age, race, tumor type, ER, PR, HER2, nodes, and primary locations have significantly influenced the overall survival rate (OS) and the specific survival rate (BCSS) for breast cancer in contrast to mastectomy. Chemotherapy significantly lowers the death rate in both. We observed that married patients receiving chemotherapy have significantly improved survival, highlighting the potential impact of marriage on prognosis.

A study suggests that breast cancer is more aggressive in young patient. Conversely, middle-aged patients tend to exhibit better outcomes compared to younger and older counterparts [12]. Multiple studies have indicated that women under 45 have lower survival rates and higher risk of recurrence than older women. However, a comparative study between patients under 40 years old and middle-aged women found that the former group had better survival prospects, except those diagnosed with stage III [13]. Although around 30-40% of breast cancer patient are over 70 years old, managing elder breast cancer remains a controversy due to the lack of clinical trial data (enrollment rate is less than 20%) [14, 15]. Furthermore, male patients have been found to have higher mortality rates across all stages of breast cancer compared to their female counterparts [16].

The impact of therapy on survival in the context of cancer treatment is a complex topic, and it can vary depending on several factors such as:

- Cancer stages: determine the extent and spread of cancer within the body. As the cancer progresses and spreads to other parts of the body, it is classified as stages I through IV [4, 17, 18, 19]. The 5-year survival rate for cases diagnosed with local, regional, and distant cancer is reported to be 99%, 85%, and 27%, respectively [20, 21].
- The TNM system [22] employs alphanumeric codes to characterize tumor size, lymph node, and metastasis of cancer, which vary depending on the specific cancer type.
- The grade of cancer [23] describes the microscopic appearance of cancer cells. Lower grades signify slower-growing cancer, while higher grades indicate faster growth.
- Hormone status [24] : Certain breast cancers are influenced by female naturally hormones such as estrogen and progesterone. Breast cancer cells possess receptors on their surface that bind to these hormones.
- Treatment: Chemotherapy utilizes drugs to target and destroy cancer cells throughout the body, while radiation therapy uses high-energy beams to target and destroy cancer cells in specific areas. The combination of these treatments aims to attack cancer from multiple angles and improve overall outcomes [14, 15].
- The sequence of radiation with surgery [25] : the sequence of radiation with surgery can have an impact on survival. It is based on individual patient characteristics and the type and stage of the cancer being treated.

Most Previous research that studied the impact of breast cancer treatment focused on analyzing single treatment options such as 'chemotherapy only' [13, 26, 27, 28], 'Radiotherapy-only' [2, 29, 30, 31], 'both chemotherapy and hormone therapy' [32] or 'both chemotherapy and radiotherapy' [33]. In contrast, our current research explores the effects of four treatment approaches: chemotherapy only', 'radiotherapy-only', 'both chemotherapy and radiotherapy' and 'no therapy'. Furthermore, while most previous studies have conducted a stratified analysis based on a single variable such as stage [34], risk groups [23, 35, 36], or age [12, 5, 16], our research takes a more comprehensive approach by performing stratified analysis using 15 variables: grade, stage, age, breast subtype, ER status, HER2 status, laterality, marital status, metastasis status, nodal status, PR status, race, radiation sequence with surgery, sex, and tumor size. By considering a broader range of variables, our research aims to provide a more nuanced understanding of the impact of treatment on different breast cancer patients.

2. Methodology

The study focused on BCSS, measuring the time from breast cancer diagnosis to breast cancer-specific death, treating other deaths as censored data. OS, a secondary outcome, tracked the time from diagnosis to death or last follow-up, treating lost-to-follow-up as censored data [15, 19]. In our statistical analysis, survival curves were generated using the Kaplan-Meier method, and the log-rank test was used to determine the statistical significance of differences between groups in BCSS or OS rates between the survival curves. We considered a P-value of less than 0.05 as statistically significant. We used Cox proportional hazards regression model to identify factors (treatment, tumor characteristics, and patient demographics) that influence OS and BCSS. It estimates hazard ratios (HRs) with 95% CIs to quantify how the risk of death or disease progression changes based on different factors, considering censored data and accounting for the timing of events. The significant variables from the univariate analysis were included in the multivariate analysis. All statistical analysis was performed using IBM SPSS Statistics, version 25 [37] .

2.1. Survival analysis [24, 38]

Survival analysis studies time-to-event data using Kaplan-Meier curves and the log-rank test. The Kaplan-Meier estimator calculates the survival function St for different time intervals. The survival function is estimated as:

$$S(t) = \prod_{i=1}^{n}(1 - \frac{d_i}{n_i}) \tag{1}$$

where

- S(t) is the estimated survival probability at time t.
- n_i is the number of individuals at risk (alive) at the previous time point
- d_i is the number of observed events (deaths) that occur at time t.

The log-rank test compares survival curves to determine significant differences between groups. The test statistic X^2 is calculated as:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{2}$$

Where:

- O is the observed number of events (death) in each group.
- E is the expected number of events in each group, assuming no survival difference.

2.2. Machine learning

The research studied the effectiveness of ten different algorithms to identify the most suitable method for predicting OS and BCSS outcome, including Adaptive Boosting (AdaBoost), C5.0 Decision Trees, Gradient Boosting Machines (GBM), Linear Discriminant Analysis (LDA), Multilayer Perceptron (MLP), Naive Bayes (NB), Neural Networks (NN), Random Forest (RF), Recursive Partitioning and Regression Trees (RPART), and Bagged tree (treebag). We evaluated algorithm performance with: accuracy, Root Mean Squared Error (RMSE), and Receiver Operating Characteristic (ROC).

2.3. Data sources and data cleaning

The study used the November 2019 SEER cancer database [39] via SEER*Stat 8.3.8. We excluded patients with non-breast cancer's cause of death and those with 'blank' or 'unknown' variables. We focused on 'active follow-up' patients to mitigate censored data issues. We used data collected after 2010, resulting in 142,902 patients with 21 variables: including patient characteristics (age, race, sex, marital status), treatment regimens (chemotherapy status, radiotherapy status, radiation sequence with surgery), tumor characteristics (size, grade,

AJCC stage, metastasis status, ER status, PR status, HER2 status, nodal status, laterality, breast subtype) Our goal is to understand treatment and patient characteristics' impact on breast cancer survival and prognosis.

3. Results

3.1. Demographics and key variables

The study included a total of 142,902 patients, with 20,129 patients in the chemotherapy only' group, 50,034 in the radiotherapy only' group, 30,889 receiving 'both chemotherapy and radiotherapy, and 41,850 not receiving chemotherapy or radiotherapy treatment. The baseline characteristics of the different treatment groups are summarized in Table 1, providing an overview of the patient demographics and key variables.

Table 1. Baseline characteristics of patients with different treatment groups

Characteristics	**None (41850)**	**Chemo only (20129)**	**Radio only (50034)**	**Both (30889)**	**Total (142902)**	**p-value**
Age at diagnosis						
<70	21228(50.7%)	16726(83.1%)	32320(64.6%)	26562(86%)	96836(67.76%)	< 0.001
70+	20622(49.3%)	3403(16.9%)	17714(35.4%)	4327(14%)	46066(32.24%)	
Breast Subtype						
Her2+/HR+	2647(19.73%)	3982(29.69%)	1649(12.29%)	5135(38.28%)	13413(9.39%)	< 0.001
Her2+/HR-	1101(18.43%)	2091(35.01%)	414(6.93%)	2367(39.63%)	5973(4.18%)	
Her2-/HR+	35356(32.57%)	9858(9.08%)	46130(42.5%)	17202(15.85%)	108546(75.96%)	
Triple Negative	2746(18.34%)	4198(28.04%)	1841(12.3%)	6185(41.32%)	14970(10.48%)	
Metastasis status						
M0	39996(95.6%)	18267(90.7%)	49134(98.2%)	29785(96.4%)	137182(96%)	< 0.001
M1	1854(4.4%)	1862(9.3%)	900(1.8%)	1104(3.6%)	5720(4%)	
Nodal status						
N0	33221(79.4%)	10119(50.3%)	43579(87.1%)	13312(43.1%)	100231(70.14%)	< 0.001
N1	6607(15.8%)	7059(35.1%)	5312(10.6%)	11178(36.2%)	30156(21.1%)	
N2	963(2.3%)	1459(7.2%)	627(1.3%)	3878(12.6%)	6927(4.85%)	
N3	1059(2.5%)	1492(7.4%)	516(1%)	2521(8.2%)	5588(3.91%)	
Chemotherapy						
No	41850(100%)	0	50034(100%)	0	91884(64.3%)	< 0.001
Yes	0	20129(100%)	0	30889(100%)	51018(35.7%)	
ER Status						
Positive	37734(90.2%)	13471(66.9%)	47614(95.2%)	21814(70.6%)	120633(84.42%)	< 0.001
Negative	4116(9.8%)	6658(33.1%)	2420(4.8%)	9075(29.4%)	22269(15.58%)	
HER2 Status						
Positive	3748(9%)	6073(30.2%)	2063(4.1%)	7502(24.3%)	19386(13.57%)	< 0.001
Negative	38102(91%)	14056(69.8%)	47971(95.9%)	23387(75.7%)	123516(86.43%)	
Laterality						
Right	20394(48.7%)	9773(48.6%)	24740(49.4%)	15294(49.5%)	70201(49.13%)	0.027

Left	21456(51.3%)	10356(51.4%)	25294(50.6%)	15595(50.5%)	72701(50.87%)	
Marital status						
Single	5577(13.3%)	3139(15.6%)	5841(11.7%)	4356(14.1%)	18913(13.23%)	< 0.001
Married	20604(49.2%)	11832(58.8%)	29585(59.1%)	19121(61.9%)	81142(56.78%)	
Other	15669(37.4%)	5158(25.6%)	14608(29.2%)	7412(24%)	42847(29.98%)	
PR Status						
Positive	33506(80.1%)	10604(52.7%)	43216(86.4%)	17630(57.1%)	104956(73.45%)	< 0.001
Negative	8344(19.9%)	9525(47.3%)	6818(13.6%)	13259(42.9%)	37946(26.55%)	
Race						
White	34287(81.9%)	15483(76.9%)	42336(84.6%)	24028(77.8%)	116134(81.27%)	< 0.001
Black	3968(9.5%)	2637(13.1%)	3966(7.9%)	4375(14.2%)	14946(10.46%)	
Other	3595(8.6%)	2009(10%)	3732(7.5%)	2486(8%)	11822(8.27%)	
Radiation						
No	41850(100%)	20129(100%)	0(0%)	0(0%)	61979(43.37%)	< 0.001
Yes	0(0%)	0(0%)	50034(100%)	30889(100%)	80923(56.63%)	
Radiation sequence with surgery						
No radiation	41850(66.38%)	20126(31.92%)	587(0.93%)	482(0.76%)	63045(44.12%)	< 0.001
Before surgery	0(0%)	0(0%)	122(34.76%)	229(65.24%)	351(0.25%)	
After surgery	0(0%)	3(0%)	48486(61.8%)	29971(38.2%)	78460(54.9%)	
Both	0(0%)	0(0%)	36(26.67%)	99(73.33%)	135(0.09%)	
Intraoperative	0(0%)	0(0%)	803(88.14%)	108(11.86%)	911(0.64%)	
Sex						
Male	596(1.4%)	246(1.2%)	133(0.3%)	207(0.7%)	1182(0.83%)	< 0.001
Female	41254(98.6%)	19883(98.8%)	49901(99.7%)	30682(99.3%)	141720(99.17%)	
Tumor Size						
< 50	38683(92.4%)	16903(84%)	48274(96.5%)	25944(84%)	129804(90.83%)	< 0.001
$50+$	3167(7.6%)	3226(16%)	1760(3.5%)	4945(16%)	13098(9.17%)	
Stage						
I	31693(75.7%)	9452(47%)	42892(85.7%)	12671(41%)	96708(67.67%)	< 0.001
II	8001(19.1%)	8182(40.6%)	6116(12.2%)	15841(51.3%)	38140(26.69%)	
IV	2156(5.2%)	2495(12.4%)	1026(2.1%)	2377(7.7%)	8054(5.64%)	
Grade						
I	12364(29.5%)	1480(7.4%)	18917(37.8%)	2756(8.9%)	35517(24.85%)	< 0.001
II	20280(48.5%)	7634(37.9%)	24458(48.9%)	11879(38.5%)	64251(44.96%)	
III	9097(21.7%)	10926(54.3%)	6597(13.2%)	16126(52.2%)	42746(29.91%)	
IV	109(0.3%)	89(0.4%)	62(0.1%)	128(0.4%)	388(0.27%)	
Year of diagnosis						
2010	7781(18.6%)	3902(19.4%)	9125(18.2%)	5908(19.1%)	26716(18.7%)	< 0.001
2011	8014(19.1%)	3987(19.8%)	9671(19.3%)	6200(20.1%)	27872(19.5%)	
2012	8382(20%)	3957(19.7%)	10043(20.1%)	6299(20.4%)	28681(20.07%)	
2013	9005(21.5%)	4319(21.5%)	10435(20.9%)	5991(19.4%)	29750(20.82%)	
2014	8668(20.7%)	3964(19.7%)	10760(21.5%)	6491(21%)	29883(20.91%)	
BCSS						
Alive	39511(94.4%)	18333(91.1%)	49233(98.4%)	29291(94.8%)	136368(95.43%)	< 0.001
Dead	2339(5.6%)	1796(8.9%)	801(1.6%)	1598(5.2%)	6534(4.57%)	

Overall survival					
Alive	36888(88.1%)	17840(88.6%)	48244(96.4%)	28916(93.6%)	131888(92.29%)< 0.001
Dead	4962(11.9%)	2289(11.4%)	1790(3.6%)	1973(6.4%)	11014(7.71%)

3.2. Comparison of survival between different groups

Based on the information provided, univariate and multivariate analyses were conducted to identify prognostic factors that could predict OS and BCSS in the cohort. In the univariate analysis Table 2, all variables except year of diagnosis were found to significantly impact on OS and BCSS, so they were included in the multivariate analysis.

Table 2. Univariate analysis for BCSS and OS in all patients

Variables	OS		BCSS	
	HRs (95% CI)	P	HRs (95% CI)	P
Age at diagnosis				
<70	Reference		Reference	
70+	2.57 (2.47-2.67)	<0.001	1.63 (1.56-1.72)	<0.001
Breast Subtype				
Her2+/HR+	Reference		Reference	
Her2+/HR-	1.42 (1.29-1.56)	<0.001	1.74 (1.55-1.95)	<0.001
Her2-/HR+	0.744 (0.70-0.79)	<0.001	0.60 (0.55-0.65)	<0.001
Triple Negative	2.05 (1.91-2.21)	<0.001	2.50 (2.29-2.73)	<0.001
Breast Subtype				
M0	Reference		Reference	
M1	12.46 (11.93-13.01)	<0.001	22.47 (21.37-23.64)	<0.001
Nodal status				
N0	Reference		Reference	
N1	2.08 (1.90-2.18)	<0.001	3.60 (3.39-3.82)	<0.001
N2	3.29 (3.08-3.50)	<0.001	6.36 (5.88-6.88)	<0.001
N3	6.91 (6.54-7.30)	<0.001	14.68 (13.72-15.7)	<0.001
Chemotherapy				
Yes	Reference		Reference	
No	1.11 (1.07-1.15)	<0.001	1.90 (1.81-1.997)	<0.001
ER Status				
Negative	Reference		Reference	
Positive	2.45 (2.35-2.55)	<0.001	3.63 (3.45-3.81)	<0.001
PR Status				
Negative	Reference		Reference	
Positive	2.24 (2.16-2.33)	<0.001	3.44 (3.28-3.61)	<0.001
HER 2				
Negative	Reference		Reference	
Positive	0.80 (0.76-0.84)	<0.001	0.668 (0.627-0.711)	<0.001
Laterality				
Right	Reference		Reference	

Continued on next page

Left	1.05 (1.01-1.09)	0.01	1.02 (0.97-1.07)	0.461
Marital status				
Single	Reference		Reference	
Married	0.55 (0.52-0.58)	<0.001	0.52 (0.48-0.56)	<0.001
other	1.25 (1.18-1.32)	<0.001	1.01 (0.94-1.08)	0.787
Race				
White	Reference		Reference	
Black	1.63 (1.55-1.72)	<0.001	1.92 (1.80-2.05)	<0.001
Other	0.71 (0.66-0.77)	<0.001	0.76 (0.69-0.85)	<0.001
Radiation				
No	Reference		Reference	
Yes	0.37 (0.36-0.39)	<0.001	0.42 (0.39-0.44)	<0.001
Radiation sequence with surgery				
No radiation	Reference		Reference	
Before surgery	2.48 (2.02-3.03)	<0.001	3.803 (3.080-4.696)	<0.001
After surgery	0.294 (0.282-0.307)	<0.001	0.291 (0.276-0.307)	<0.001
Both	0.802 (0.466-1.383)	0.43	1.353 (0.785-2.331)	0.277
Intraoperative	0.1 (0.05-0.200)	<0.001	0.041 (0.010-0.165)	<0.001
Sex				
Male	Reference		Reference	
Female	0.58 (0.494-0.682)	<0.001	0.875 (0.677-1.131)	0.309
Tumor Size				
<50	Reference		Reference	
50+	4.541 (4.36-4.73)	<0.001	7.215 (6.865-7.584)	<0.001
Stage				
I	Reference		Reference	
II	2.27 (2.17-2.38)	<0.001	4.424 (4.134-4.734)	<0.001
IV	13.85 (13.23-14.50)	<0.001	37.23 (34.88-39.74)	<0.001
Grade				
I	Reference		Reference	
II	1.205 (0.163-1.257)	<0.001	1.079 (0.060-1.104)	<0.001
III	2.344 (1.275-2.430)	<0.001	1.236 (0.182-1.307)	<0.001
IV	4.703 (2.563-4.878)	<0.001	2.702 (0.542-2.910)	0.008
Year of diagnosis				
2010	Reference		Reference	
2011	1.067 (0.960-1.186)	0.23	1.042 (0.915-1.187)	0.531
2012	1.061 (0.955-1.179)	0.27	1.038 (0.912-1.182)	0.573
2013	1.068 (0.961-1.188)	0.22	1.024 (0.898-1.168)	0.72
2014	1.105 (0.991-1.233)	0.07	1.096 (0.958-1.253)	0.183
chemo_rad				
None	Reference		Reference	
Chemo_only	0.936(0.891-0.983)	0.009	1.560(1.467-1.658)	<0.001
Rad_only	0.280(0.265-0.296)	<0.001	0.267(0.246-0.289)	<0.001
Both	0.492(0.467-0.519)	<0.001	0.849(0.797-0.905)	<0.001

The result of multivariate analysis shown in Table 3 revealed a better survival in patients received radiotherapy, according to both OS and BCSS (HRs=0.280, 95% CI=0.265-0.296, p<0.001; HRs= 0.267, 95% CI=0.246- 0.289, p<0.001) followed by patients who receive a combination of both chemotherapy and radiotherapy, (HRs= 0.492, 95% CI=0.467-0.519, p<0.001; HRs=0.849, 95% CI=0.797-0.905, p<0.001).

Table 3. Multivariate analysis for BCSS and OS in all patients

Variables	OS		BCSS	
	HRs (95% CI)	P	HRs (95% CI)	P
Age at diagnosis				
<70	Reference		Reference	
70+	1.092 (0.863-1.971)	**0.003**	1.853 (1.776-1.935)	**<0.001**
Breast Subtype				
Her2+/HR+	Reference		Reference	
Her2+/HR-	1.445 (1.189-1.757)	**<0.001**	1.092 (0.925-1.288)	0.299
Her2-/HR+	0.794 (0.72-0.875)	**<0.001**	0.532 (0.489-0.58)	**<0.001**
Triple Negative	1.457 (1.202-1.765)	**<0.001**	1.161 (0.985-1.369)	0.076
Metastasis status				
M0	Reference		Reference	
M1	1.784 (1.608-1.98)	**<0.001**	1.266 (1.153-1.39)	**<0.001**
Nodal status				
N0	Reference		Reference	
N1	2.273 (2.169-2.381)	**<0.001**	3.499 (3.284-3.728)	**<0.001**
N2	4.197 (3.922-4.491)	**<0.001**	6.819 (6.271-7.414)	**<0.001**
N3	8.078 (7.619-8.564)	**<0.001**	14.633 (13.616-15.726)	**<0.001**
Chemotherapy				
Yes	Reference		Reference	
No	1.975 (1.878-2.078)	**<0.001**	1.154 (1.108-1.201)	**<0.001**
ER Status				
Negative	Reference		Reference	
Positive	0.519 (0.436-0.619)	**<0.001**	0.476 (0.409-0.553)	**<0.001**
PR Status				
Negative	Reference		Reference	
Positive	0.749 (0.698-0.804)	**<0.001**	0.688 (0.651-0.727)	**<0.001**
HER2				
Positive	Reference		Reference	
Negative	0.8 (0.76-0.84)	**<0.001**	0.907 (0.86-0.956)	**<0.001**
Laterality				
Right	Reference		Reference	
Left	0.977 (0.93-1.025)	0.342	1.062 (1.023-1.103)	**0.002**
Marital status				
Single	Reference		Reference	
Married	0.96 (0.893-1.032)	0.264	0.825 (0.779-0.874)	**<0.001**

Continued on next page

other	0.955 (0.902-1.011)	0.111	1.082 (1.022-1.145)	**0.007**
Race				
White	Reference		Reference	
Black	1.102 (0.992-1.225)	0.072	1.086 (1.03-1.146)	**0.002**
Other	1.153 (1.025-1.298)	**0.018**	0.848 (0.781-0.921)	**<0.001**
Radiation				
No	Reference		Reference	
Yes	0.4 (0.38-0.42)	**<0.001**	0.368 (0.354-0.383)	**<0.001**
Radiation sequence with surgery				
No radiation	Reference		Reference	
Before surgery	0.726 (0.577- 0.913)	**<0.001**	0.590 (0.474-0.734)	**<0.001**
After surgery	0.676 (0.605-0.756)	**<0.001**	0.531 (0.481-0.588)	**<0.001**
Both	0.708 (0.407-1.231)	0.221	0.458 (0.264-0.796)	0.006
Intraoperative	0.584 (0.145-2.347)	0.449	0.301 (0.150-0.607)	0.001
Sex				
Male	Reference		Reference	
Female	0.77 (0.595-0.998)	**0.048**	1.319 (1.12-1.553)	**0.001**
Tumor Size				
<50	Reference		Reference	
50+	1.221 (1.156-1.291)	**<0.001**	1.167 (1.113-1.224)	**<0.001**
Stage				
I	Reference		Reference	
II	1.899 (1.716-2.101)	**<0.001**	1.514 (1.398-1.639)	**<0.001**
IV	2.225 (1.944-2.546)	**<0.001**	2.023 (1.805-2.269)	**<0.001**
Grade				
I	Reference		Reference	
II	1.34 (1.201-1.496)	**<0.001**	1.07 (1.005-1.139)	**0.035**
III	1.758 (1.574-1.964)	**<0.001**	1.228 (1.149-1.313)	**<0.001**
IV	1.564 (1.182-2.069)	**0.002**	1.413 (1.122-1.78)	**0.003**

Table 4 presents a comparison of OS and BCSS between different treatment patients stratified by grade. Additionally, Kaplan-Meier survival curves for the effect of chemotherapy on grade for BCSS in Fig. 2 and for OS in Fig. 1 are shown. The key findings from this analysis are as follows:

- All breast cancer grades can benefit from different treatments according to OS and BCSS, except for grade I and grade II patients in BCSS who do not benefit from 'chemotherapy only' treatment but can benefit from other treatments. The highest treatment that can benefit grade I, grade II and grade IV patients is 'Radiotherapy-only' treatment, according to both OS and BCSS.
- For patients in grade III, 'both chemotherapy and radiotherapy' treatment provide the highest survival according to OS, while 'Radiotherapy-only' treatment provides the highest survival according to BCSS.

These results highlight the importance of considering the grade of breast cancer when determining the most effective treatment approach. The findings emphasize that different treatments can yield favorable outcomes based on the grade of the disease, and personalized treatment strategies can optimize survival rates for patients in different grades of breast cancer.

The study investigated the impact of various treatments on breast cancer patients at different stages. Table 5 compared outcomes, helping identify optimal treatments. Kaplan-Meier curves for OS in Fig. 3 and Fig. 4 illustrated the effect of different treatments on BCSS stratified by stage.

Table 4. Comparison of BCSS and OS between different treatments in a specific grade

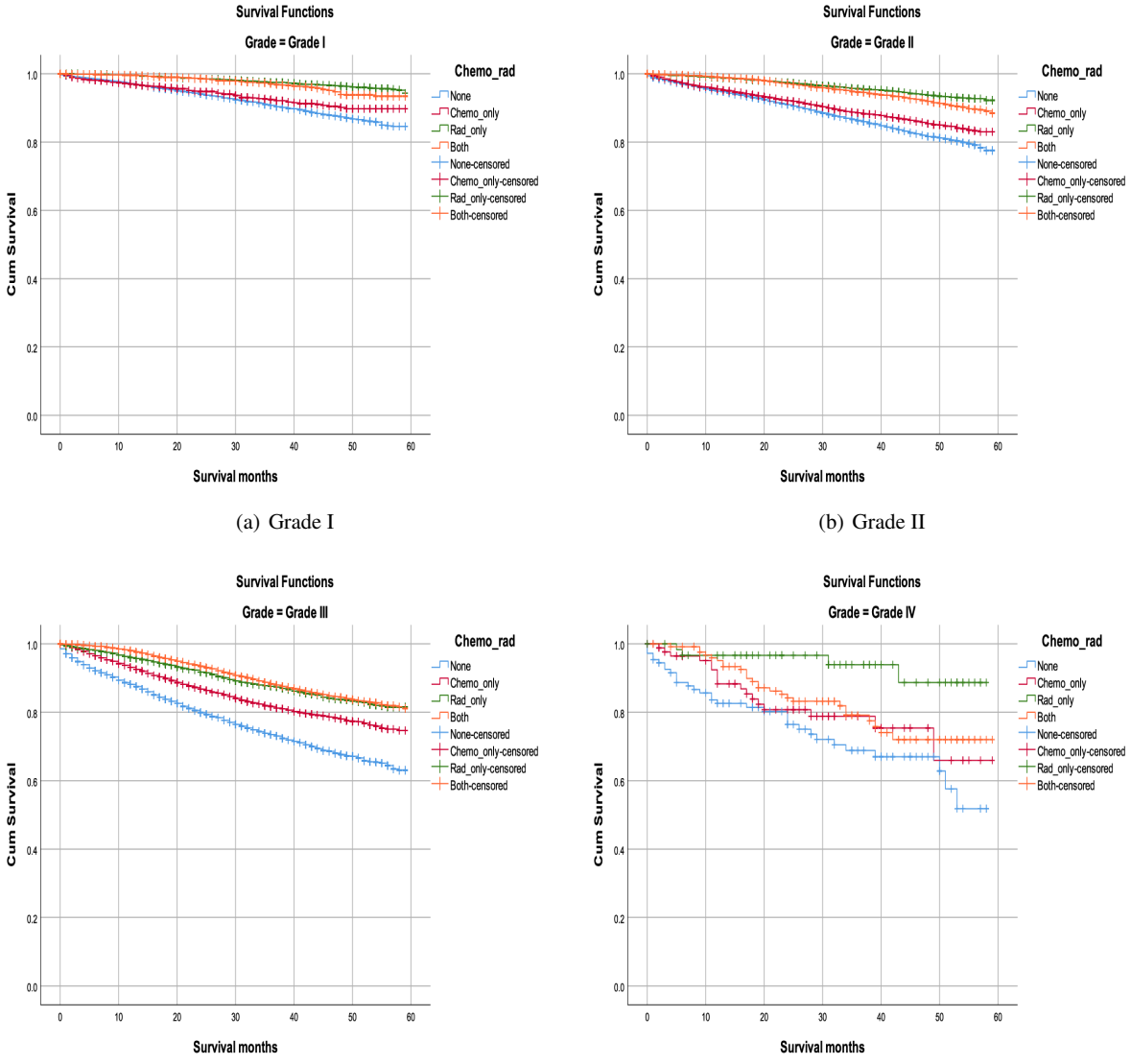
Grade	OS			BCSS		
	#Event	HRs (95% CI)	P	#Event	HRs (95% CI)	P
Grade I (35517)						
None	870	Reference		229	Reference	
Chemo_only	91	0.790 (0.636-0.980)	**0.032**	52	1.727 (1.278-2.334)	**<0.001**
Rad_only	353	0.249 (0.220-0.282)	**<0.001**	65	0.175 (0.1330-0.230)	**<0.001**
Both	74	0.336 (0.265-0.426)	**<0.001**	46	0.799 (0.582-1.097)	0.165
Grade II (64251)						
None	2136	Reference		886	Reference	
Chemo_only	682	0.810 (0.743-0.883)	**<0.001**	479	1.375 (1.230-1.536)	**<0.001**
Rad_only	804	0.292 (0.269-0.317)	**<0.001**	343	0.302 (0.266-0.342)	**<0.001**
Both	508	0.366 (0.332-0.403)	**<0.001**	382	0.667 (0.592-0.752)	**<0.001**
Grade III (42746)						
None	1924	Reference		1206	Reference	
Chemo_only	1499	0.618 (0.578-0.661)	**<0.001**	1251	0.824 (0.761-0.891)	**<0.001**
Rad_only	629	0.408 (0.372-0.446)	**<0.001**	390	0.404 (0.360-0.452)	**<0.001**
Both	1365	0.355 (0.331-0.380)	**<0.001**	1147	0.476 (0.439-0.517)	**<0.001**
Grade IV (388)						
None	32	Reference		18	Reference	
Chemo_only	17	0.693 (0.385-1.249)	0.223	14	0.997 (0.496-2.006)	0.994
Rad_only	4	0.190 (0.067-0.536)	**0.002**	3	0.259 (0.076-0.879)	**0.03**
Both	26	0.595 (0.354-0.998)	**0.049**	23	0.276 (0.083-0.918)	**0.036**

Overall, patients across all breast cancer stages can benefit from various treatments, as both OS and BCSS indicate. However, the findings highlight the importance of tailoring treatment strategies based on the specific stage of breast cancer. 'Radiotherapy-only' treatment appears to be the optimal choice for stage I patients, while 'both chemotherapy and radiotherapy' treatment benefits stage II and stage IV patients. The results emphasize the need for personalized treatment approaches to optimize survival outcomes for breast cancer patients at different stages.

The study employed a more comprehensive approach by conducting a stratified analysis for 15 variables. By comparing OS and BCSS among patients receiving different treatments across all variables, we aimed to identify the optimal treatment strategy for each variable category in table 6.

There seem to be disparities between the outcomes of the BCSS and OS examinations concerning the optimal treatment choices for various groups of patients. While 'both therapies' are typically linked to improved survival rates in BCSS, distinct interventions can yield favorable survival outcomes in OS, contingent upon the patient subset. There are some discrepancies between the treatments that yield the highest survival rates according to the BCSS and OS measures, as well as different subgroups of patients. Here is a summary of the findings:

- Age: for patients under 70 years old, chemotherapy only' and 'both treatments have the highest survival rates according to BCSS. However, according to OS, 'Radiotherapy-only' and 'both' treatments yield better survival rates for all ages.
- Breast Subtypes: The highest survival rates for patients with different breast subtypes are observed in Her2-/HR+ and triple-negative subtypes when they receive 'both' treatments, according to BCSS. However, OS suggests that different treatments provide high survival rates across various subtypes.



(a) Grade I

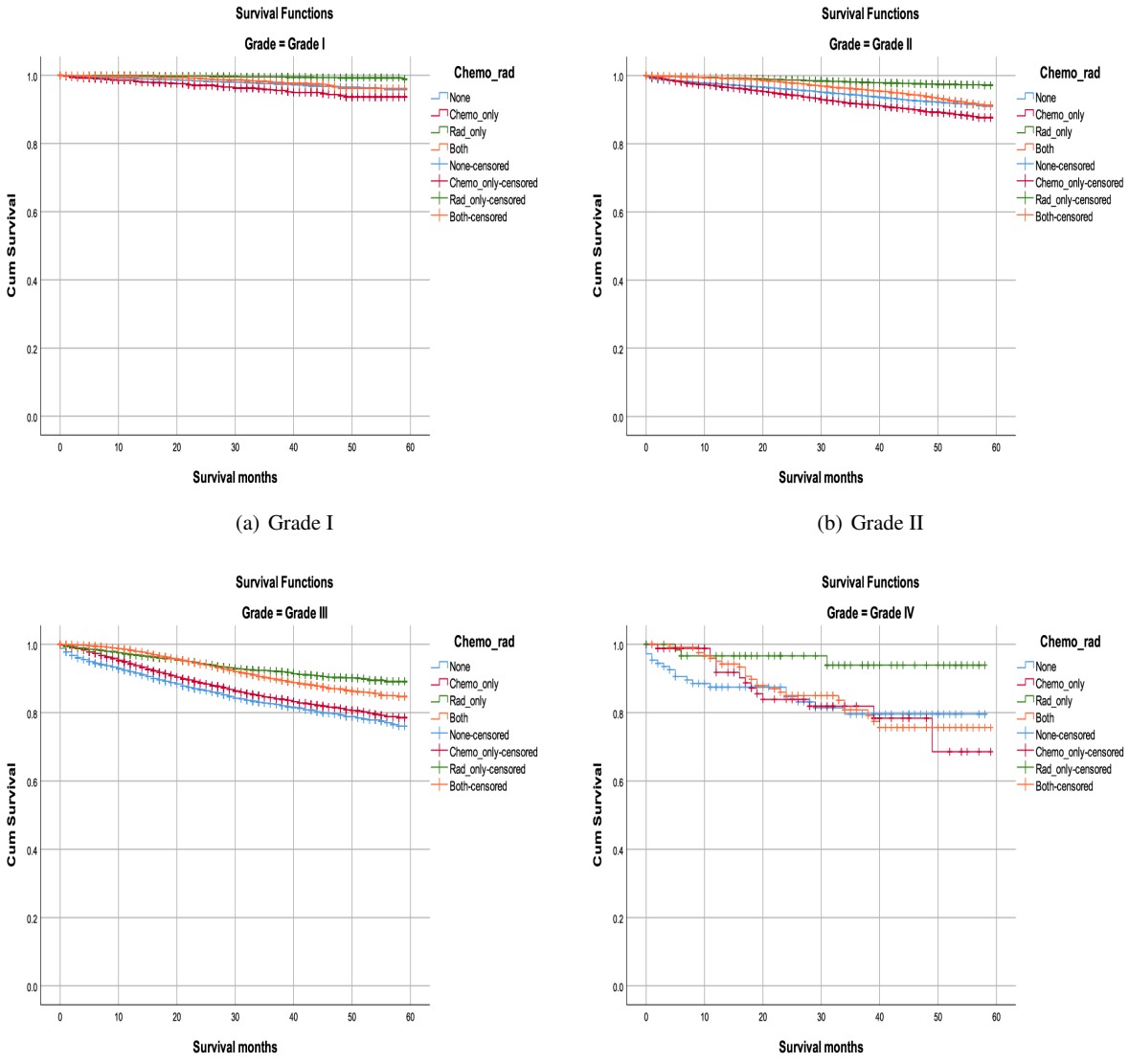
(b) Grade II

(c) Grade III

(d) Grade IV

Figure 1. Kaplan–Meier survival curves for the effect of different treatments stratified by grade for OS

- Metastasis Status: BCSS indicates that patients with MO metastasis status who receive 'both' treatments have the best survival rates. On the other hand, OS suggests that different treatments provide high survival rates regardless of metastasis status.
- Nodal Status: BCSS shows that patients with NO and N1 nodal status who receive both' treatments have the highest survival rates. Conversely, OS indicates that different treatments provide high survival rates for patients with different nodal statuses.
- ER Status: BCSS suggests that patients with different ER statuses who receive 'both' treatments have better survival rates. In contrast, OS indicates that different treatments provide high survival rates across different ER statuses.

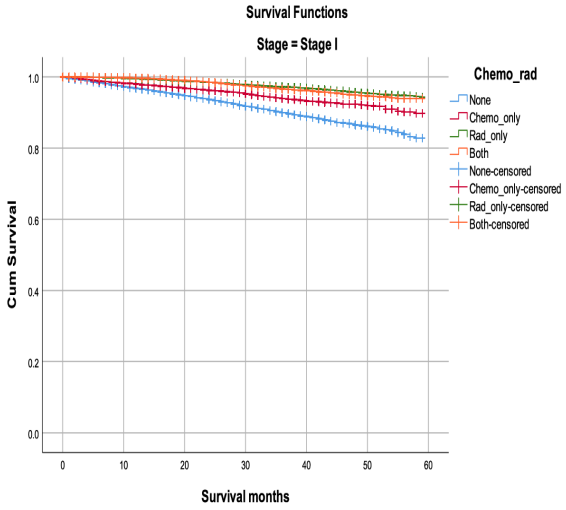


(a) Grade I (b) Grade II

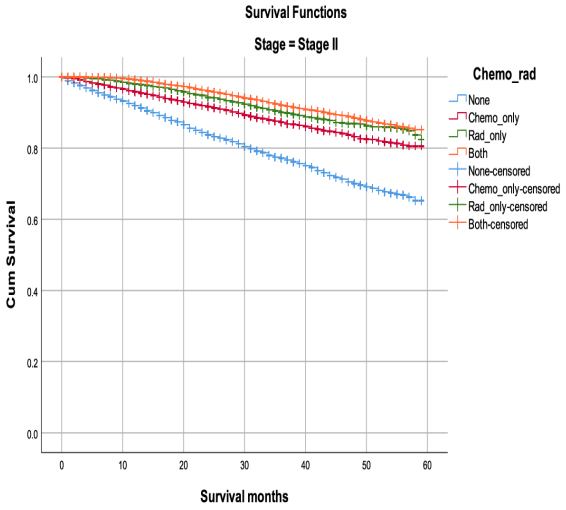
(c) Grade III (d) Grade IV

Figure 2. Kaplan–Meier survival curves for the effect of different treatments stratified by grade for BCSS

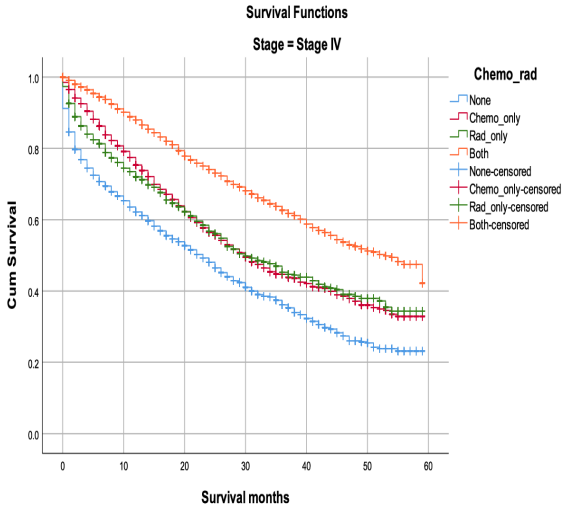
- Grade: BCSS suggests that patients with different grades who receive 'both' treatments have better survival rates. However, OS indicates that different treatments provide high survival rates across different grades.
- HER2 Status: Negative HER2 patients who receive both treatments have better survival rates according to BCSS. OS suggests that different treatments provide high survival rates for most negative HER2 patients, except for those who receive 'chemotherapy only.'
- Laterality: According to BCSS, both 'chemotherapy only' and 'both' treatments provide better survival for both right and left breast laterality patients. However, according to OS, different treatments provide high survival rates regardless of breast laterality.
- Marital Status: BCSS indicates that 'both treatments provide better survival rates for all marital statuses. In addition, for 'married' patients, chemotherapy only' also yields high survival rates. On the other hand,



(a) Stage I



(b) Stage II

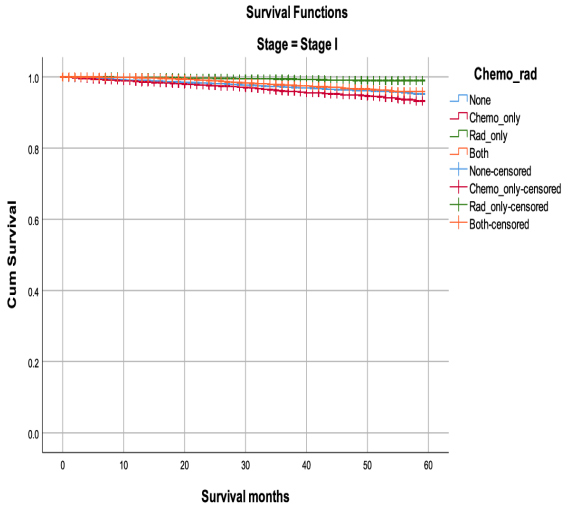


(c) Stage IV

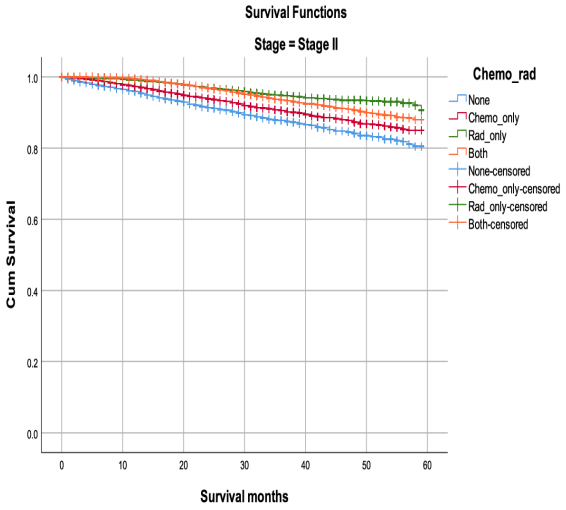
Figure 3. Kaplan–Meier survival curves for the effect of different treatments stratified by stage for OS

according to OS, 'Radiotherapy-only' and 'both' treatments result in the highest survival rates for both 'single' and 'married' patients. All treatments provide better survival rates for 'other' marital status patients.

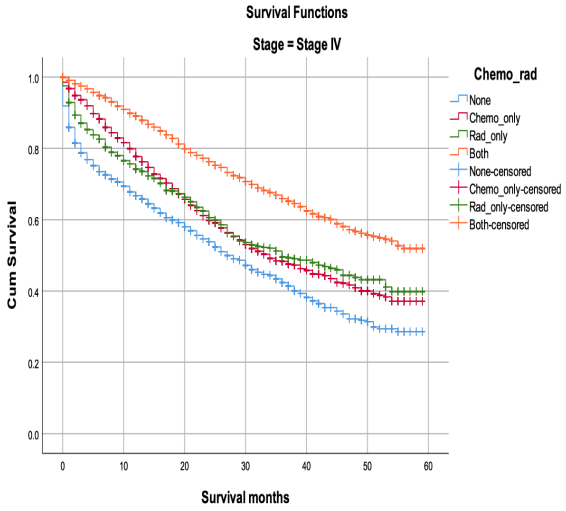
- PR Status: BCSS suggests that 'both' treatments provide better survival rates for all PR statuses. In contrast, according to OS, different treatments provide high survival rates across different PR statuses. Race: According to BCSS, 'both treatments provide better survival rates for both 'white' and 'black' patients. For 'other' patients, 'chemotherapy only' and 'both' treatments yield the highest survival rates. However, according to OS, different treatments provide high survival rates for most ethnicities, except for 'black' patients who receive chemotherapy only.



(a) Stage I



(b) Stage II



(c) Stage IV

Figure 4. Kaplan–Meier survival curves for the effect of different treatments stratified by stage for BCSS

- Sex: BCSS indicates that 'chemotherapy only' and 'both' treatments result in the highest survival rates for 'male patients. For 'female' patients, receiving 'both' treatments leads to the highest survival rates. Conversely, according to OS, different treatments provide high survival rates for both sexes.
- Tumor Size: According to BCSS, patients with tumor sizes less than 50 mm who receive 'both' treatments have the highest survival rates. However, patients with tumor sizes larger than 50 mm have lower survival rates regardless of the treatment received. In contrast, OS suggests different treatments provide high survival rates across different tumor sizes.

Table 5. Comparison of BCSS and OS between different treatments in a specific stage

Stage	Treatment	OS			BCSS		
		#Event	HRs (95% CI)	P	#Event	HRs (95% CI)	P
Stage I							
	None	31693	Reference		2406	Reference	
	Chemo_only	9452	0.580 (0.523-0.643)	< 0.001	431	1.372 (1.191-1.580)	< 0.001
	Rad_only	42892	0.266 (0.246-0.287)	< 0.001	917	0.207 (0.176-0.243)	< 0.001
	Both	12671	0.299 (0.265-0.336)	< 0.001	309	0.731 (0.624-0.855)	< 0.001
Stage II							
	None	8001	Reference		1424	Reference	
	Chemo_only	8182	0.508 (0.466-0.553)	< 0.001	814	0.726 (0.651-0.810)	< 0.001
	Rad_only	6116	0.361 (0.324-0.403)	< 0.001	424	0.356 (0.305-0.415)	< 0.001
	Both	15841	0.292 (0.269-0.317)	< 0.001	946	0.461 (0.417-0.511)	< 0.001
Stage IV							
	None	2156	Reference		1132	Reference	
	Chemo_only	2495	0.674 (0.619-0.733)	< 0.001	1044	0.702 (0.641-0.767)	< 0.001
	Rad_only	1026	0.699 (0.627-0.780)	< 0.001	449	0.723 (0.643-0.812)	< 0.001
	Both	2377	0.390 (0.355-0.429)	< 0.001	718	0.404 (0.366-0.447)	< 0.001

Table 6. Comparison of BCSS and OS between different treatment patients for all other variables

Variables	Treatments	BCSS HRs (95% CI)	BCSS P	OS HRs (95% CI)	OS P
Age at diagnosis					
<70	**None**	Reference		Reference	
	Chemo_only	0.826 (0.756-0.903)	**<0.001**	1.639 (1.523-1.765)	**<0.001**
	Rad_only	1.735 (1.607-1.874)	**<0.001**	0.331 (0.301-0.363)	**<0.001**
	Both	0.225 (0.2-0.254)	**<0.001**	0.921 (0.854-0.992)	**0.03**
<70	**None**	Reference		Reference	
	Chemo_only	1.105 (0.979-1.248)	0.107	1.095 (1.008-1.190)	**0.031**
	Rad_only	2.074 (1.798-2.392)	**<0.001**	0.319 (0.299-0.341)	**<0.001**
	Both	0.345 (0.299-0.399)	**<0.001**	0.516 (0.467-0.570)	**<0.001**
Breast Subtype					
Her2+/HR+	**None**	Reference		Reference	
	Chemo_only	3.548 (2.88-4.37)	**<0.001**	0.473 (0.410-0.545)	**<0.001**
	Rad_only	2.38 (1.936-2.926)	**<0.001**	0.334 (0.270-0.413)	**<0.001**
	Both	1.316 (0.976-1.773)	0.071	0.220 (0.187-0.259)	**<0.001**
Her2+/HR-	**None**	Reference		Reference	
	Chemo_only	2.385 (1.918-2.965)	**<0.001**	0.497 (0.413-0.598)	**<0.001**
	Rad_only	1.267 (1.022-1.571)	**0.031**	0.487 (0.353-0.673)	**<0.001**
	Both	0.674 (0.472-0.960)	0.028	0.410 (0.308-0.547)	**<0.001**
Triple Negative	**None**	Reference		Reference	
	Chemo_only	2.248 (1.825-2.769)	**<0.001**	0.537 (0.464-0.621)	**<0.001**
	Rad_only	0.704 (0.570-0.871)	**0.001**	0.317 (0.259-0.388)	**<0.001**
	Both	0.532 (0.367-0.772)	**0.001**	0.356 (0.284-0.446)	**<0.001**

3.3. Machine-learning-based outcome prediction

Based on the results for both BCSS and OS classification using various algorithms shown in table 7:

Table 7. Model performance for BCSS and OS

Algorithm	BCSS				OS			
	Accuracy	Sensitivity	RMSE	ROC_score	Accuracy	Sensitivity	RMSE	ROC_score
AdaBoost	0.98	0.99	0.14	0.9	96.87	0.71	0.16	0.95
C5.0	0.98	0.99	0.14	0.91	0.98	0.99	0.15	0.88
GBM	0.98	0.99	0.14	0.9	0.98	0.76	0.15	0.87
LDA	0.97	0.97	0.18	0.98	0.97	0.64	0.17	0.82
MLP	0.95	1	0.21	0.52	0.92	0	0.28	0.5
NB	0.95	0.6	0.2	0.99	0.94	0.6	0.25	0.78
NN	0.98	0.99	0.14	0.89	0.97	0.6	0.18	0.79
RF	0.98	0.78	0.14	0.87	0.98	0.78	0.12	0.86
RPART	0.98	0.8	0.15	0.89	0.97	1	0.16	0.86
Treebag	0.98	0.99	0.14	0.9	0.97	0.78	0.16	0.88

BCSS Classification:

- C5.0, GBM, and AdaBoost excel in accuracy, sensitivity, and ROC scores. C5.0 slightly outperforms.
- NB has a top ROC score (0.99) but lower sensitivity, impacting positive instance identification.
- LDA achieves a good ROC score (0.98) but slightly lower accuracy.
- Treebag, RPART, and RF have high accuracy but lower sensitivity and ROC scores.
- MLP excels in sensitivity (1.00) but has lower accuracy and ROC score (0.52).

OS Classification:

- C5.0, RF, and GBM offer high accuracy, reasonable sensitivity, low RMSE, and good ROC scores.
- Treebag has good discrimination ability with slightly lower accuracy and sensitivity.
- RPART boasts perfect sensitivity but slightly lower accuracy and ROC score.
- LDA, NN, and AdaBoost have lower sensitivity, potentially leading to more false negatives.
- NB has the lowest accuracy and sensitivity.
- MLP struggles with identifying positive instances effectively (0.00 sensitivity).

4. Discussion

This study suggests that there are differences in the recommended treatment approaches based on patient's individual circumstances. This study may help healthcare professionals to assess the potential outcomes and plan appropriate treatment strategies for patients with different characteristics. Differences in individual patient characteristics contribute to variations in the optimal treatment for breast cancer patients, resulting in distinct recommendations for OS and BCSS. The results underscore the importance of personalized treatment strategies for breast cancer patients, taking into account factors such as:

- Age: Age plays a role in cancer patients' treatment and survival rates. Patients older than 70 tend to have worse survival rates than younger patients. A combination of chemotherapy and radiotherapy' is significantly reducing the risk of death, as reflected by lower HRS in both BCSS and OS with highly significant p-values.
- Subtypes: patients with HR+ and HER2- subtypes generally have a more favorable prognosis compared to those with HR- and HER2+ subtypes. Supplemental therapies are recommended to improve treatment outcomes. Triple-negative and negative hormone receptor (ER/PR) subtypes have a worse prognosis.

Treatment for triple-negative breast cancer often involves a combination of chemotherapy and radiation therapy, similar to HER2- subtype patients.

- Metastasis and nodal status profoundly influence survival. Patients without metastasis (MO) generally experience improved outcomes, with treatments, especially chemotherapy, reducing the risk of death. Furthermore, the number of affected lymph nodes (nodal status) plays a pivotal role, with lower nodal stages (NO and N1) associated with better treatment responses. These results underscore the importance of early detection and intervention to prevent the progression of the disease to advanced stages. Patients with N2 and N3 nodal status may need supplemental therapy.
- Laterality: The breast cancer patients with tumors in the right breast tend to have better survival rates compared to those with tumors in the left breast. However, it's important to note that the difference in survival rates based on laterality is not consistently observed in all studies and may vary among different patient populations.
- Marital status: All marital status appears to influence the effectiveness of a combination both chemotherapy and radiotherapy.
- Race: It is important to clarify that medical treatment recommendations should not be based solely on a patient's race or ethnicity. The results show that white patients generally experiencing better survival outcomes compared to Black patients. Regardless of the race, both chemotherapy and radiation therapy are recommended for breast cancer patients.
- The sequence of radiation with surgery: The sequence of radiation with surgery plays a significant role in determining patient survival rates. Current study suggests that administering radiation after surgery, utilizing intraoperative radiation, or employing a combination of radiation before and after surgery generally to leads to better survival outcomes than radiation before or without radiation. However, the optimal treatment approach should be determined by considering the patient's specific characteristics and their cancer, as well as consulting with healthcare professionals.
- Sex: Female patients have advantage in survival rates than males in breast cancer. Male patients may benefit from chemotherapy only' or a combination of chemotherapy and radiotherapy, while female patients may benefit from the combination of chemotherapy and radiotherapy.
- Tumor size: Tumor size is one of the key factors considered in treatment decision-making for breast cancer. Larger tumors generally have a higher risk of spreading to lymph nodes or distant sites, and may require more aggressive treatment approaches. If the tumor size is less than 50 mm, 'both' therapy is recommended. For tumors larger than 50 mm, a supplemental therapy is recommended. The term 'supplemental therapy' is broad and can include additional treatments such as targeted therapy, hormonal therapy, or extended adjuvant therapy.
- According BCSS, the combination of 'both' chemotherapy and radiotherapy is the recommended treatment for the following patients: HER-, triple negative, stage II, grade III, HER2-/HR+ and tumor size ¡ 50. 'Radiotherapy' is the recommended treatment for patients with stage I, all grades except grade III.
- According to OS, radiotherapy only or in combination with chemotherapy is recommended for majority cases.

The results of BCSS classification using the accuracy, sensitivity, RMSE, and ROC scores for each algorithm show that:

- The majority of algorithms, including C5.0, GBM, AdaBoost, NN, Treebag, and RPART have relatively high accuracy, sensitivity, and ROC scores, indicating strong overall performance and effectiveness in accurately classifying patients into BCSS categories. Additionally, the low RMSE values suggest an accurate prediction of survival durations. These algorithms show promise for BCSS prediction. However, the RPART and RF models show lower sensitivity compared to the others.
- They achieve a high accuracy of 0.98 or above. This indicates that these models are capable of making correct predictions for a large proportion of the instances.
- Most models, such as C5.0, GBM, AdaBoost, NN, Treebag, exhibit a high sensitivity score of 0.99. This means that these models are effective in correctly identifying positive instances, as they have a low false negative rate.

- However, both RPART and RF models have lower sensitivity scores compared to the other models, with values of 0.80 and 0.78, respectively. This suggests that these models may struggle to accurately detect positive instances, resulting in a higher false negative rate.
- The RMSE values are relatively consistent across all models, ranging from 0.14 to 0.21. These values indicate the average difference between the predicted values and the actual values. Lower RMSE values generally indicate better predictive performance.
- ROC scores measure the overall classification performance of the models. Most models, including C5.0, GBM, AdaBoost, NN, Treebag, and RPART, achieve relatively high ROC scores between 0.89 and 0.91. These scores indicate that these models have a good ability to distinguish between positive and negative instances.
- Notably, the MLP model has a lower ROC score of 0.52, suggesting that it may struggle with classification and distinguishing between positive and negative instances.

The results for OS classification using accuracy, sensitivity, RMSE, and ROC scores for each algorithm:

- C5.0, RF, and GBM performed well in terms of accuracy, achieving scores above 0.97. They also showed competitive performance in sensitivity and ROC scores, indicating their ability to accurately classify patients into OS categories. Additionally, the RMSE values suggest a relatively accurate prediction of survival durations. These algorithms demonstrate promise for OS prediction.
- C5.0 appears to be the best-performing model overall. It achieves high accuracy (0.98) and sensitivity (0.99), indicating that it has a low rate of both false positives and false negatives. Additionally, it has a reasonably high ROC score (0.88), indicating good discrimination between positive and negative instances.
- RF and GBM come next in the ranking. While their accuracy (0.98) is comparable to C5.0, their sensitivity values are slightly lower, indicating a higher rate of false negatives. However, they still demonstrate good overall performance.
- Treebag, RPART and AdaBoost also perform well with high accuracy values (0.97) and reasonably good sensitivity scores. They have good ROC scores, suggesting relatively good discrimination capabilities. The high ROC score suggests excellent discrimination ability.
- LDA and NN achieved a high accuracy score, but NN has a lower sensitivity score compared to LDA. However, the lower sensitivity scores compared to the top-performing algorithms indicate potential limitations in identifying patients with adverse OS outcomes and discriminating between different OS categories. The low ROC score indicates fair discrimination ability.
- On the lower end, models like NB and MLP show comparatively lower performance. They have lower sensitivity scores and lower ROC scores, indicating a higher rate of false negatives and poor discrimination between positive and negative instances.

5. Conclusion

The research addresses the following objectives: (1) Determine the effect of radiotherapy and chemotherapy on breast cancer survival by analyzing the survival curves and hazard ratios; (2) Identify factors associated with improved survival outcomes, including tumor characteristics, patient demographics, and treatment regimens; (3) Develop and compare predictive models using statistics and machine learning algorithms to accurately estimate survival probabilities based on treatment variables. Based on the findings of the stratified analysis considering OS and BCSS, it can be concluded that the optimal treatment for breast cancer patients varies based on several factors, including age, breast subtype, metastasis status, nodal status, ER/PR status, laterality, marital status, sex, tumor size, and the sequence of radiation with surgery. The 'both' treatment, which combines chemotherapy and radiation therapy, generally emerges as the most effective treatment option, consistently demonstrating higher survival rates across many analyzed variables. However, there are certain subgroups where alternative approaches may be more beneficial. In terms of BCSS, patients with specific criteria such as HER-, triple-negative, stage II, grade III, HER2-/HR+, and tumor size ¡ 50 benefited most from a combined chemotherapy and radiotherapy

approach, while those in stage I, with grades other than III found 'radiotherapy only' to be adequate. In the context of OS analysis, Radiotherapy only' or 'in combination with chemotherapy emerged as more effective treatments across a wide range of cases, often outperforming 'chemotherapy only.' Machine learning models were developed to forecast OS and BCSS, and the C5.0 algorithm consistently demonstrated robust overall performance. These discoveries enhance the decision-making process for breast cancer treatment.

6. Future Work:

Future research in studying breast cancer treatment should consider several important aspects: Combined Variable Analysis: It is crucial to conduct combined variable analysis, which takes into account multiple factors simultaneously. This approach provides a more comprehensive understanding of the complex interactions between variables and treatment outcomes. By considering various factors together, more precise treatment strategies can be identified. Exploring Immunotherapy and Targeted Therapy: The investigation of emerging treatment modalities, such as immunotherapy and targeted therapy, is of utmost importance. These therapies have demonstrated promising results in various cancer types, and assessing their effectiveness specifically in breast cancer patients can yield valuable insights for improving treatment outcomes. Evaluating long-term side effects and quality of life: Understanding the long-term side effects associated with different treatments is essential. It is also crucial to assess the impact of these side effects on patients' quality of life. By evaluating these factors, we can gain a better understanding of the overall treatment experience and make informed decisions that prioritize both efficacy and patients' well-being. By addressing these research areas, we can enhance our understanding of breast cancer treatment, improve patient outcomes, and make strides towards reducing the burden of this disease.

Statements and Declarations

Authors' contributions

AE, MA, and BT contributed to the conception; design; and development of the methodology; AE and MA contributed to the acquisition of data and analysis; and wrote, reviewed, and revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was not supported by any grant funding.

Availability of data and materials

The dataset generated and analyzed during the current study is available in the Surveillance, Epidemiology, and End Results (SEER) database. The URL of the database is https://seer.cancer.gov/

Ethics approval and consent to participate

Considering SEER database is publicly available. We signed a Data-Use Agreement for the SEER 1973–2019 Research Data File to get access conditions. Also, we signed a Data Use Agreement for SEER Radiation Therapy and Chemotherapy Information Data extraction and usage have been approved by SEER Program.

Consent for publication

All authors approve the manuscript for publication.

Competing interests

The authors have declared that no competing interests exist.

REFERENCES

1. Melina Arnold, Eileen Morgan, Harriet Rumgay, Allini Mafra, Deependra Singh, Mathieu Laversanne, Jerome Vignat, Julie R Gralow, Fatima Cardoso, Sabine Siesling, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66:15–23, 2022.
2. Ross Lawrenson, Chunhuan Lao, Ahmed Ali, and Ian Campbell. Impact of radiotherapy on cardiovascular health of women with breast cancer. *Journal of Medical Imaging and Radiation Oncology*, 63(2):250–256, 2019.
3. Jun Wang, Wen-Wen Zhang, Chen-Lu Lian, Jia-Yuan Sun, Zhen-Yu He, and San-Gang Wu. The effect of post-mastectomy radiotherapy in patients with metaplastic breast cancer: an analysis of seer database. *Frontiers in Oncology*, 9:747, 2019.
4. Manojna Konda and Rashmi Verma. Cancer screening and prevention: Sex and gender evidence in lung, breast, and colorectal cancer. In *How Sex and Gender Impact Clinical Practice*, pages 75–99. Elsevier, 2021.
5. Hai-long Chen, Mei-qi Zhou, Wei Tian, Ke-xin Meng, and Hai-fei He. Effect of age on breast cancer patient prognoses: a population-based study using the seer 18 database. *PloS one*, 11(10):e0165409, 2016.
6. Carol E DeSantis, Jiemin Ma, Mia M Gaudet, Lisa A Newman, Kimberly D Miller, Ann Goding Sauer, Ahmedin Jemal, and Rebecca L Siegel. Breast cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(6):438–451, 2019.
7. Samta Rani, Tanvir Ahmad, and Sarfaraz Masood. Comparative analysis of breast and prostate cancer prediction using machine learning techniques. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 1*, pages 643–650. Springer, 2022.
8. Jiali Ji, Shushu Yuan, Jiawei He, Hong Liu, Lei Yang, and Xuexin He. Breast-conserving therapy is associated with better survival than mastectomy in early-stage breast cancer: A propensity score analysis. *Cancer Medicine*, 11(7):1646–1658, 2022.
9. Akimitsu Yamada, Hiraku Kumamaru, Chikako Shimizu, Naruto Taira, Kanako Nakayama, Mika Miyashita, Naoko Honma, Hiroaki Miyata, Itaru Endo, Shigehira Saji, et al. Systemic therapy and prognosis of older patients with stage ii/iii breast cancer: a large-scale analysis of the japanese breast cancer registry. *European Journal of Cancer*, 154:157–166, 2021.
10. Fei Wang, Xiang Shu, Ingrid Meszoely, Tuya Pal, Ingrid A Mayer, Zhigang Yu, Wei Zheng, Christina E Bailey, and Xiao-Ou Shu. Overall mortality after diagnosis of breast cancer in men vs women. *JAMA oncology*, 5(11):1589–1596, 2019.
11. Sonia Kaindal and B Venkataramana. Comparative effect of chemotherapy and non-chemotherapy treatments on the survival of breast cancer patients at different stages: A seer cross-sectional based study. 2024.
12. Walburga Y Joko-Fru, Adalberto Miranda-Filho, Isabelle Soerjomataram, Marcel Egue, Marie-Therese Akele-Akpo, Guy N'da, Mathewos Assefa, Nathan Buziba, Anne Korir, Bakarou Kamate, et al. Breast cancer survival in sub-saharan africa by age, stage at diagnosis and human development index: A population-based registry study. *International journal of cancer*, 146(5):1208–1218, 2020.
13. Kaiyan Huang, Jie Zhang, Yushuai Yu, Yuxiang Lin, and Chuangui Song. The impact of chemotherapy and survival prediction by machine learning in early elderly triple negative breast cancer (etnbc): a population based study from the seer database. *BMC geriatrics*, 22(1):268, 2022.
14. Jingtong Zhai, Yun Wu, Fei Ma, Virginia Kaklamani, and Binghe Xu. Advances in medical treatment of breast cancer in 2022. *Cancer Innovation*, 2(1):1–17, 2023.
15. Kathryn P Trayes and Sarah EH Cokenakes. Breast cancer treatment. *American family physician*, 104(2):171–178, 2021.
16. Helen M Johnson, William Irish, Mahvish Muzaffar, Nasreen A Vohra, and Jan H Wong. Quantifying the relationship between age at diagnosis and breast cancer-specific mortality. *Breast cancer research and treatment*, 177:713–722, 2019.
17. Armando E Giuliano, Stephen B Edge, and Gabriel N Hortobagyi. of the ajcc cancer staging manual: breast cancer. *Annals of surgical oncology*, 25:1783–1785, 2018.
18. Steven S Coughlin. Social determinants of breast cancer risk, stage, and survival. *Breast cancer research and treatment*, 177:537–548, 2019.
19. Marco Alfonse, Mostafa M Aref, and Abdel-Badeeh M Salem. An ontology-based system for cancer diseases knowledge management. *International Journal of Information Engineering and Electronic Business*, 6(6):55–63, 2014.
20. Angela N Giaquinto, Hyuna Sung, Kimberly D Miller, Joan L Kramer, Lisa A Newman, Adair Minihan, Ahmedin Jemal, and Rebecca L Siegel. Breast cancer statistics, 2022. *CA: a cancer journal for clinicians*, 72(6):524–541, 2022.
21. Muhammad Akram, Mehwish Iqbal, Muhammad Daniyal, and Asmat Ullah Khan. Awareness and current knowledge of breast cancer. *Biological research*, 50:1–23, 2017.
22. Eithar Seifeldeen Mohammed Mustafa et al. *Developing the Classification of Breast Cancer System Using Random Forest Technique*. PhD thesis, Sudan University of Science & Technology, 2022.
23. Fiorenza De Rose, Bruno Meduri, Maria Carmen De Santis, Antonella Ferro, Lorenza Marino, Riccardo Ray Colciago, Fabiana Gregucci, Valentina Vanoni, Giovanni Apolone, Serena Di Cosimo, et al. Rethinking breast cancer follow-up based on individual risk and recurrence management. *Cancer Treatment Reviews*, 109:102434, 2022.
24. Ahmed Barakat, Aaina Mittal, David Ricketts, and Benedict A Rogers. Understanding survival analysis: Actuarial life tables and the kaplan–meier plot. *British Journal of Hospital Medicine*, 80(11):642–646, 2019.
25. Omolola Salako, Alero A Roberts, Victor I Isibor, Oluwatimilehin Babatunde, Omolara Fatiregun, and Chukwumere N Nwogu. Innovative breast cancer awareness and advocacy campaign. *Journal of global oncology*, 3(2):169–176, 2017.
26. Lea Rossi, Denise Stevens, Jean-Yves Pierga, Florence Lerebours, Fabien Reyal, Mathieu Robain, Bernard Asselain, and Roman Rouzier. Impact of adjuvant chemotherapy on breast cancer survival: a real-world population. *PloS one*, 10(7):e0132853, 2015.
27. SE Ward, GR Holmes, A Ring, PD Richards, JL Morgan, JW Broggio, K Collins, MWR Reed, and L Wyld. Adjuvant chemotherapy for breast cancer in older women: an analysis of retrospective english cancer registration data. *Clinical Oncology*, 31(7):444–452, 2019.
28. Benjamin M Parsons, Dipesh Uprety, Angela L Smith, Andrew J Borgert, and Leah L Dietrich. A us registry–based assessment of use and impact of chemotherapy in stage i her2-positive breast cancer. *Journal of the National Comprehensive Cancer Network*, 16(11):1311–1320, 2018.

29. Lauren T Steward, Feng Gao, Marie A Taylor, and Julie A Margenthaler. Impact of radiation therapy on survival in patients with triple-negative breast cancer. *Oncology letters*, 7(2):548–552, 2014.
30. Minoru Miyashita, Onyinye B Balogun, Olufunmilayo I Olopade, and Dezheng Huo. The optimization of postoperative radiotherapy in de novo stage iv breast cancer: evidence from real-world data to personalize treatment decisions. *Scientific Reports*, 13(1):2880, 2023.
31. Jizhao Wang, Yuchen Sun, Jingkun Qu, Huang Zuo, Xixi Zhao, Lin Liu, Jinteng Feng, Jiansheng Wang, and Guangjian Zhang. Survival analysis for male ductal and lobular breast cancer patients with different stages. *Future Oncology*, 15(2):167–180, 2019.
32. Zohreh Sanaat, Jalil Vaez Gharamaleki, Yosra Vaez Gharamaleki, and Roya Dolatkhah. The analysis of breast cancer survival in east azerbaijan province of iran; prognostic impact of chemotherapy and hormone therapy protocols. *Cancer Treatment and Research Communications*, 32:100578, 2022.
33. Xinrong Li, Jin Zhang, Chenxiao Ye, Junquan Zhu, Kaibo Guo, and Yong Guo. Impact of chemotherapy and radiotherapy on the survival of elderly esophageal cancer patients undergoing surgery: a seer database analysis. *BMC gastroenterology*, 21:1–11, 2021.
34. Nai-Chen Shih, Pei-Tseng Kung, Wei-Yin Kuo, and Wen-Chen Tsai. Association of treatment delay and stage with mortality in breast cancer: a nationwide cohort study in taiwan. *Scientific Reports*, 12(1):18915, 2022.
35. Zheng Wang, Hui Wang, Xi Sun, Yan Fang, Shuang-Shuang Lu, Shu-Ning Ding, Xiao-Song Chen, and Kun-Wei Shen. A risk stratification model for predicting overall survival and surgical benefit in triple-negative breast cancer patients with de novo distant metastasis. *Frontiers in Oncology*, 10:14, 2020.
36. Xiaozhu Liu, Song Yue, Haodong Huang, Minjie Duan, Binyi Zhao, Jin Liu, and Tianyu Xiang. Risk stratification model for predicting the overall survival of elderly triple-negative breast cancer patients: a population-based study. *Frontiers in Medicine*, 8:705515, 2021.
37. IBM Corp. *IBM SPSS Statistics for Windows*, 2017. Version 25.0. Armonk, NY: IBM Corp.
38. John P Klein, Hans C van Houwelingen, Joseph G Ibrahim, and Thomas H Scheike. *Handbook of Survival Analysis*. CRC Press, 2016.
39. Surveillance Research Program, National Cancer Institute. *SEER*Stat software*, 2020. Version 8.3.5, Accessed 2nd January 2020. Available at: www.seer.cancer.gov/seerstat.