Journal of Mathematics Research; Vol. 15, No. 5; October 2023
ISSN 1916-9795 E-ISSN 1916-9809
Published by Canadian Center of Science and Education

On the Implications of Ignoring Competing Risk in Survival Analysis: The Case of the Product-Limit Estimator

Joseph Acquah[1], Senyefia Bosson-Amedenu[1] & Eric Adubuah[2]

[1] Department of Mathematical Sciences, University of Mines and Technology, Tarkwa, Ghana

[2] Department of Economics Education, University of Education, Winneba, Ghana

Correspondence: Senyefia Bosson-Amedenu, Department of Mathematical Sciences, University of Mines and Technology, P.O. Box 237, Tarkwa, Ghana. E-mail: senyefiaa.bosson@hcce.edu.gh

Received: June 4, 2023 Accepted: August 30, 2023 Online Published: October 16, 2023

doi:10.5539/jmr.v15n5p1 URL: https://doi.org/10.5539/jmr.v15n5p1

Abstract

Although the Kaplan-Meier (KM) is a single event model, it is frequently used in literature with datasets that are assumed to be cause-specific without any proper verification. It is crucial to evaluate the implication of this on the probability estimates. This study compares the estimates of the cumulative incidence functions to the complement of the product-limit estimator (1-KM). The KM was found to inflate probability estimates when the dataset is unverified for competing risk. Estimates with lower standard errors and a larger area under the Receiver Operation Characteristic (ROC) curve were related to datasets verified for competing events, while estimates with datasets unverified for competing risk had lower area under the ROC curve and higher standard errors. The results support the idea that since the product-limit estimator is a cause-specific model, it naturally performs better for a single event model than in the case of several events; hence, it is necessary to verify the dataset for competing events. The findings of this study clearly suggest that before choosing a modelling strategy, one should confirm competing risks in the survival dataset.

Keywords: Kaplan-Meier, competing risk, breast cancer, survival analysis, cumulative incidence function

1. Introduction

The ability to accurately predict BC disease survival and risk is critical to enhancing overall health care. Over the last decade, survival studies have been muddled by competing risks in biological and clinical research. If appropriate methodologies are not employed to analyse the dataset, the presence of competing risks in survival data leads to incorrect estimates of probability or hazards. In survival data, competing risk prevents the occurrence of the event of interest, altering the chance of witnessing the primary event, which adds bias if traditional survival methods are utilised. The presence of Competing risks in BC data tends to make real illness incidence estimations more difficult (Swampa, 2013).

The results of a review of biomedical research that compared estimates from competing and noncompeting risk models on the same dataset gave contradictory findings. On the one hand, estimates were practically equal (competing events were uncommon in the data), while on the other hand, estimates varied dramatically (high quantum of competing risk). According to these findings, there is no set standard for determining whether to use a competing or noncompeting risk model other than comparing the estimates of competing and noncompeting risk models to see if competing events are present in the survival data (Wright et al., 2020; Usman, 2019; Perera et al., 2018; Ghavami et al., 2017; Glas et al., 2016; Sapir-Pichhadze et al., 2016). Abdel-Qadir et al. (2018) found that when KM was compared to predictions made using the Cumulative Incidence Function (CIF), it overestimated the incidence of stroke patients by 39%. Austin et al. (2017) evaluated earlier survival studies that were released in 2015 and found that around 77% of the studies that were appropriate for competing risk modeling were analysed using the product-limit estimator technique. Only about 25% of respondents used CIF to support the presence or absence of rival activity. In their study, Abdel-Qadir et al. (2018) discovered that KM overstated the incidence of stroke patients by 39% when compared to predictions using the Cumulative Incidence Function (CIF). Austin et al. (2017) assessed previous survival studies published in 2015 and discovered that the KM approach was used to analyse roughly 77% of the works that were suitable for competing risk modelling. Only approximately a quarter of respondents used CIF to justify the existence or absence of competing activities.

Age is a significant variable in clinical analyses of competing risk events. According to research, the probability of competing events increases with age, from 19% in individuals under 75 to 54% in those over 75 (Glas et al., 2016;

Arani et al., 2018). According to Chapman et al. (2014), BC patients in wealthy nations encounter conflicting dangers as they get closer to the age of 70. In the last ten years, 50 publications were assessed, and almost 70% of them showed the widespread use of competing risk approach in western studies, showing the prevalence of the competing risk in data in such high-income countries (Koller et al., 2012).

The nonparametric Kaplan-Meier (KM) technique is frequently used to assess survival probability over time. Since the estimated survival probabilities are completely devoid of presumptions about the population from which the sample of survival times was selected, the product-limit estimation method is essential for analysing the dataset. The foundation of KM is the idea of independent or non-informative censorship. This suggests that the individuals whose survival times have been censored have the same outlook as the participants who are still involved in the study. Because the patient who has died is no longer at risk of failure, it is misleading if a patient passes away from a condition unrelated to the event of interest (has a probability of zero), as the KM has no mechanism to update the risk set with respect to patients who had died from causes other than failure. The KM method overestimates the odds of survival, making it inefficient for filtering circumstances in which participants are exposed to several occurrences (competing risks).When dealing with datasets that includes competing events, the Kaplan-Meier method cannot be applied. It frequently results in flawed computations and incorrect inferences when used regardless. As a workaround, the cumulative incidence function (CIF) has been widely demonstrated in scientific literature to be the most effective (Usman et al., 2019).

2. Method

The Kaplan Meier, Log rank test, and cumulative incidence function were used to explore the need to account for the risk due to comorbid conditions. Where the complement of the product-limit estimator probability estimates exceed the estimates of the cumulative incident functions, there is a chance of observing a comorbid condition at that instant. Optimisation technique was then used to determine the optimal interval for observing significant competing risk event. These techniques make it feasible to separate competing risk data points from noncompeting risk data points, thereby allowing them to be analysed separately. The receiver operation characteristic curve was used to study the significance of each breast cancer covariate in predicting survival or risk to the disease. The R software version 4.02 was used to analyse the dataset.

2.1 The Kaplan Meier (Product-Limit Estimator)

The Kaplan Meier Method is shown in Table 1 as a product-limit estimator, and the computation process for standard errors and confidence interval estimates of survival probabilities is shown in Table 2.

Table 1. Kaplan-Meier Method

Patient ID	Time (Month)	No. Censored	No. Dead (D_n)	Number at risk (R_n)	$\frac{D_n}{R_n}$	$\hat{p}_i$	$\hat{P}_k$
i	1	a	e	n	e/n	$1-(e/n)=w$	$w=\hat{P}_{k_1}$
ii	2	b	f	$n-a-e=k$	f/k	$1-(f/k)=v$	$w*v=\hat{P}_{k_2}$
iii	3	c	g	$k-b-f=m$	g/m	$1-(g/m)=x$	$w*v*x=\hat{P}_{k_3}$
iv	4	d	h	$m-c-g=q$	h/q	$1-(h/q)=y$	$w*v*x*y=\hat{P}_{k_4}$
v	5	e	i	$n-d-h=r$	i/r	$1-(i/r)=z$	$w*v*x*y*z=\hat{P}_{k_5}$

(Sources: Collet, 2015; Kleibaum and Klein, 2015)

Table 2. Standard Errors and Confidence Interval Estimates of Survival Probabilities

$\hat{P}_{k_i}$	$\frac{D_n}{R_n(R_n-D_n)}$	$\sum \frac{D_n}{R_n(R_n-D_n)}$	$\hat{P}_k \sum \frac{D_n}{R_n(R_n-D_n)}$ (Standard Error)	$\hat{P}_{k_i} \pm 1.96 \times \left(\hat{P}_k \sum \frac{D_n}{R_n(R_n-D_n)}\right)$ (confidence interval)
$\hat{P}_{k_1}$	$\frac{e}{n(e-n)} = a_1$	a_1	$\hat{P}_{k_1} a_1$	$\hat{P}_{k_1} \pm 1.96 \times (\hat{P}_{k_1} a_1)$
$\hat{P}_{k_2}$	$\frac{f}{k(f-k)} = a_2$	$a_1 + a_2 = b_1$	$\hat{P}_{k_2} b_1$	$\hat{P}_{k_2} \pm 1.96 \times (\hat{P}_{k_2} b_1)$
$\hat{P}_{k_3}$	$\frac{g}{m(g-m)} = a_3$	$b_1 + a_3 = b_2$	$\hat{P}_{k_3} b_2$	$\hat{P}_{k_2} \pm 1.96 \times (\hat{P}_{k_2} b_1)$
$\hat{P}_{k_4}$	$\frac{h}{q(h-q)} = a_4$	$b_2 + a_4 = b_3$	$\hat{P}_{k_4} b_3$	$\hat{P}_{k_4} \pm 1.96 \times (\hat{P}_{k_4} b_3)$
$\hat{P}_{k_5}$	$\frac{i}{r(i-r)} = a_5$	$b_3 + a_5 = b_4$	$\hat{P}_{k5} b_4$	$\hat{P}_{k_5} \pm 1.96 \times (\hat{P}_{k_5} b_4)$

(Sources: Collet, 2015; Kleibaum and Klein, 2015)

2.2 Cumulative Incidence Function (CIF)

The likelihood of encountering an event of type *e* at time t_f is calculated by multiplying the estimated functions of surviving prior time periods by the cause specific hazard at t_f. Now, the cumulative incidence Function (CIF) for each type of event e at the time t_f is the summation of the incidence probabilities up to the time t_f, from $f' = 1$ to $f' = f$ expressed in Equation (1).

$$CIF_e(t_f) = \sum_{f'=1}^{f} \hat{I}_e(t_f) = \sum_{f'=1}^{f} \hat{h}_e(t_{f'}) \times \hat{S}(t_{f'-1}) \tag{1}$$

where, the estimates $\hat{h}_e(t_f)$ and $\hat{S}(t_{f-1})$, are the hazard and survival probabilities of experiencing the event type e at time t_f respectively.

All subjects will eventually experience the main event of interest if they wait long enough, according to the non-informative censoring assumption, as there aren't any competing events in this case. The cumulative incidence function for non-informative censoring (non-competing risk phenomena) is given in Equation (2),

$$CIF(t) = F(t) = \Pr(T \le t) = 1 - S(t) = 1 - KM \tag{2}$$

The cumulative incidence estimator can be expressed using the Kaplan Meier estimates as indicated in Equation (3),

$$CIF(t) = \sum_{t_i < t} \frac{d_i}{n_i} \times K(t_i) \tag{3}$$

where, t_i denotes ordered observed times, n_i is the number of patients at risk beyond t_i; d_i is the number of event of interest at t_i and $K(t_i)$ is the Product-limit estimate of probability free of all events at time t_i.

2.3 Log-Rank Test

A popular test using the null hypothesis of no survival difference between two or more independent groups is the Log-Rank Test (LRT) (survival curves). It involves statistically comparing each group's projected survival curves (using

the Product-limit estimator). The observed number of incidents in each category is compared to the anticipated number under the null hypothesis (i.e. provided the survival curves are identical). The alternative option is that, with $\alpha = 0.05$, the two survival curves are not quite identical. Equation (4) illustrates how the Log-rank test statistic is related to the chi-square test statistic.

$$\text{Log} - rankStatistic = \frac{(O_2 - E_2)^2}{Var(O_2 - E_2)} \approx \chi_1^{\ 2} \tag{4}$$

where; $e_{1j} = \left(\frac{n_{1j}}{n_{1j} + n_{2j}} \right) \times \left(m_{1j} + m_{2j} \right)$ is the expected cell count with $\left(\frac{n_{1j}}{n_{1j} + n_{2j}} \right)$ being the proportion in risk set ; where $\left(m_{1j} + m_{2j} \right)$ is the number of failures over both groups and $O_i - E_i = \sum_{j=1}^{n} \left(m_{ij} - e_{ij} \right)$ for n number of failure times. The hazard ratio is assumed to be constant throughout time in the LRT model.

2.3 Data Scope

Five hundred and fifty-eight (558) patients with BC tumor features and demographics were evaluated in this retrospective study, which spanned the years 2010 to 2015 and included diagnosed BC patients at Ghana's Korle Bu Teaching Hospital and Effia Nkwanta Regional Hospital. Korle Bu Teaching Hospital is the primary referral center for the entire southern Ghana region, as well as beyond. The information was gathered from patients' records in the hospital's archives, which included their survival status and treatment information over time. The study includes only patients with enough information across the five-year period of interest.

The interpretation of ROC curve scores have been given according to Bosson-Amedenu et al. (2022); as No discrimination (AUC=0.5); Poor discrimination (0.6≥AUC>0.5); Acceptable discrimination (0.7≥AUC>0.6); Excellent discrimination (0.8≥AUC>0.7); Outstanding discrimination (AUC>0.9).

3. Results and Discussions

Table 3 shows the demographic of patients. As seen, 54.7% of diagnoses were made in young women 50 years of age or younger, which is similar to the 54.8% discovered by Okifo et al. (2021). The average age of the BC patients in this study was 50 years old, which is consistent with the young BC patient profile projected for the area. In other studies conducted in Cape Coast (49.9 years), Kumasi (49.1 years), and Africa (50.2 years), similar outcomes were obtained (Adeloye et al., 2018). Research (Okifo et al., 2021; Ohene-Yeboah et al., 2015; Seshie et al., 2015; Boyle et al., 2012; Stark et al., 2010) supports this finding that HER2+ was the least prevalent molecular subtype (3%) and triple negative (Basal type) was the most (43%) common. About 49% of BC patients were found to be in late stages (III and IV), which is consistent with previous findings (Thomas et al., 2017). However, our study found that 46% of BC patients had Grade 3 tumors, which is roughly 4% lower than the 50% reported by others (Thomas et al., 2017). Our study's 50.4% BC metastasis is within the range of the three most prevalent sites of distant metastases of 39.8% - 55.3% found by Mensah et al. (2021). In our study, 32% of the patients had Luminal A, compared to 32.8% in another study (Mensah et al., 2021). Our study identified 85% BC lymph node metastasis, which is a little higher than the 80% reported by others (Quayson et al., 2021). The percentages of roughly 39% among Akan are similar to the 38.4% recorded by Osei-Afriyie et al. (2021) for the same ethnic group, while the 29% for GA-Adamgbe was higher than the 14.3% reported, and the 18% for Ewe was lower than the 30.9% reported by the same study.

Table 3. Patient Demographics

Characteristics	Category	No. (%)
Age	≤ 50	305 (54.7)
	> 50	253 (45.3)
Stage	I	101 (18.1)
	II	155 (27.8)
	III	211 (37.8)
	V	60 (10.8)
	Other	31(5.6)
Grade	I	179(32.1)
	II	121(21.7)
	III	258(46.2)

Hospitalisation	Yes	94 (16.8)
	No	464 (83.2)
Molecular Subtype	Luminal A	179 (32.1)
	Luminal B	121 (21.7)
	Triple Negative (Basal)	240 (43.0)
	Her2+	18 (3.2)
Recurrent	Yes	48 (8.6)
	No	510 (91.4)
Menopause	Yes	364 (65.2)
	No	194(34.8)
Ethnicity	Akan	217(38.9)
	Ga Adamgbe	131(23.5)
	Ewe	99(17.7)
	Others	111(19.9)

3.1 Accounting for Competing Risk

Figure 1 compares the estimates of three models of Cumulative Incidence Functions as explicated in Equations (1), (2) and (3) to identify ages with and without competing risk. The complement of the Product-Limit Estimator (1-KM) and cumulative incidence function estimations are compared in order to establish whether or not there is competing risk in the dataset. Any inflation of the estimates of the complement of the Product-Limit Estimator in comparison to the cumulative incidence functions would be due to the presence of competing risk in the dataset. If no failures from competing events happen, the Kaplan Meier estimates are comparable to the cumulative incidence estimates.

Indicating that there was competing risk in the dataset, Figure 1 demonstrates that Equation (2) inflated the survival probability at intervals after the 32nd month of follow-up. Except for the 60th month, when the Equation (1) estimate increased, the Equation (1) and Equation (3) projections were substantially comparable throughout the follow-up period. The peaking at the 60th month of follow-up reflects the highest incidence of competing risk. Pham et al. (2016), who found that BC patients had an increasing risk of dying from the condition but are more likely to experience competing events after five years, provide support for our results. These results support the work of Austin et al. (2017), Pham et al. (2017), and Abdel-Qadir et al. (2018).

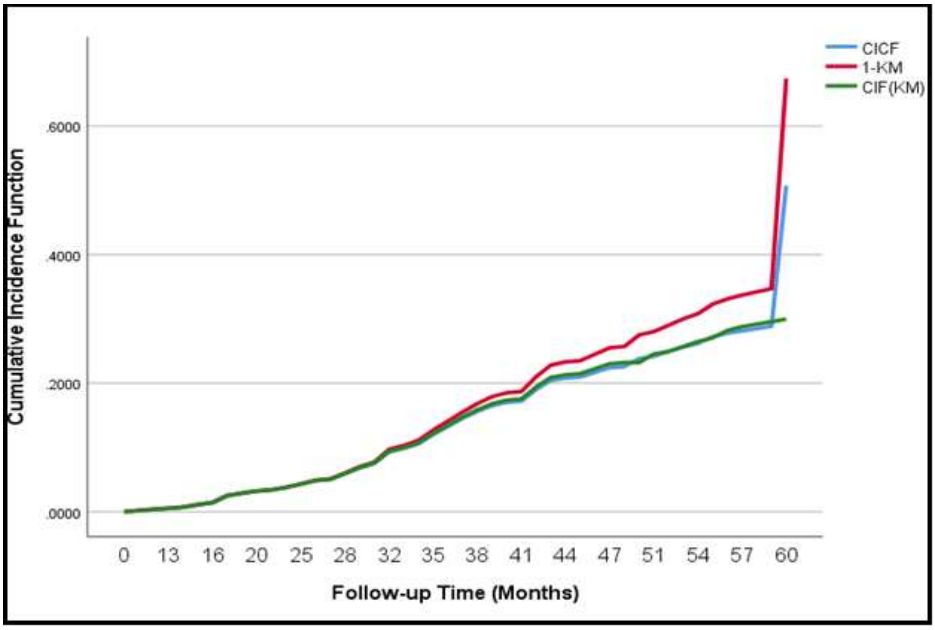

Figure 1. Comparing the Estimates of CICF, CIF (KM) and 1-KM

Figures 2 and 3 respectively illustrate the product limit estimation of survival probability for the case of BC patients with recurrent status before and after accounting for competing risk in the dataset. When competing risk is not taken into consideration, it is evident that the anticipated probability inflates. Hospitalization (Figures 4 and 5) and menopause (Figures 6 and 7) both exhibit the similar pattern. At the patient level, probability estimates can be reviewed, with ages

indicating overestimation of probabilities in favour of the 1-KM estimator being flagged as a competing risk event. The extent of the estimation, for example, one, two, or three decimal places, must be provided. However, a more relevant method to go about it is to figure out when is the best time for a patient to experience a competing risk event (comorbidities). While these numbers are available for Western countries, Africa, and specifically Ghana, has yet to provide them.

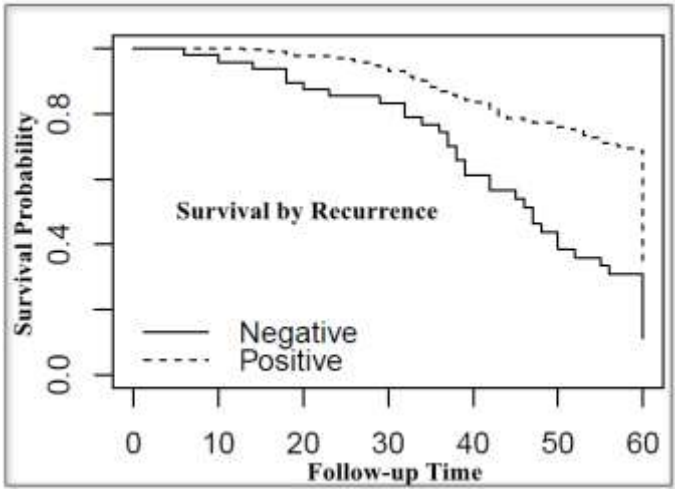

Figure 2. Survival Plot Stratified by Recurrence without accounting for Competing Risk

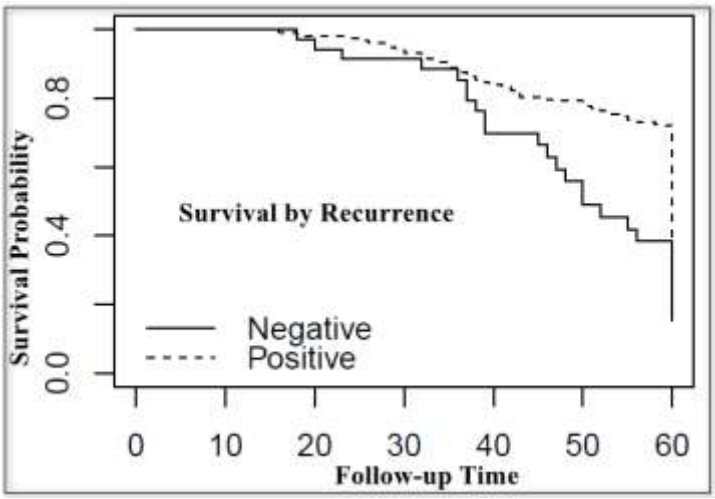

Figure 3. Survival Plot Stratified by Recurrence after accounting for Competing Risk

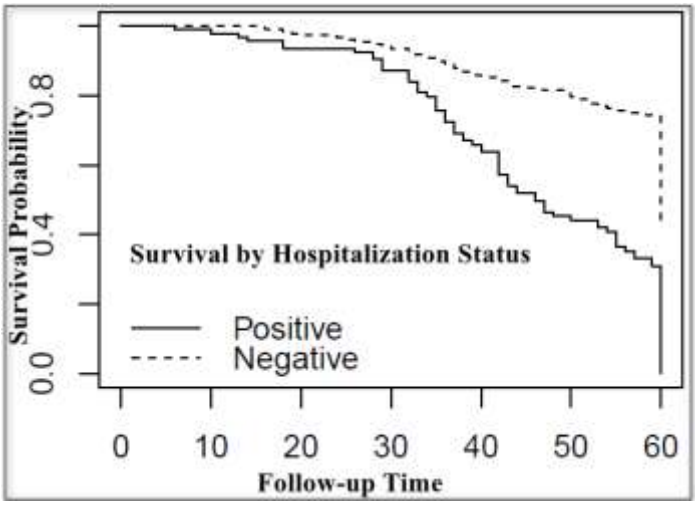

Figure 4. Survival Plot Stratified by Hospitalization without Competing Risk

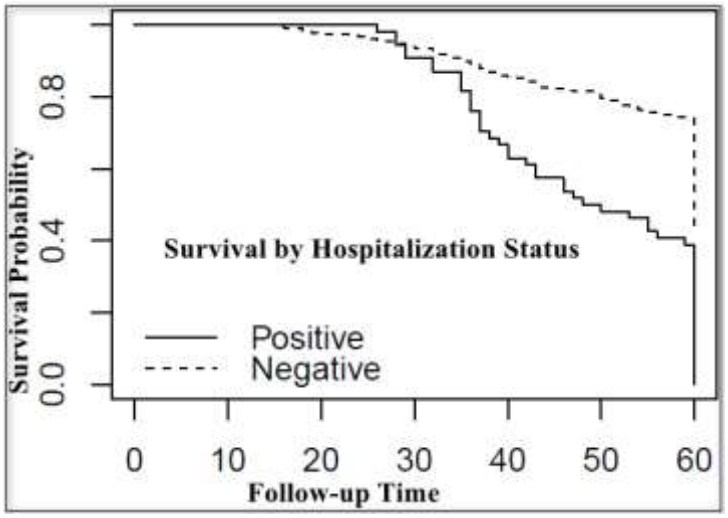

Figure 5. Survival Plot Stratified by Hospitalization after accounting for Competing Risk

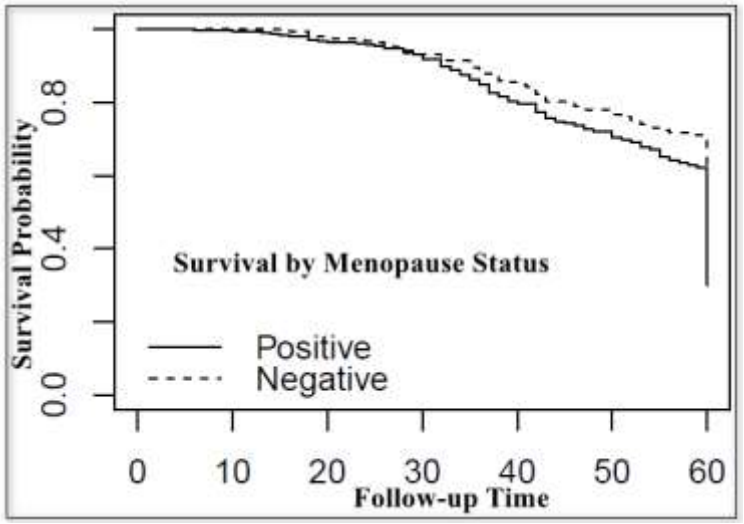

Figure 6. Survival Plot Stratified by Menopause without accounting for Competing Risk

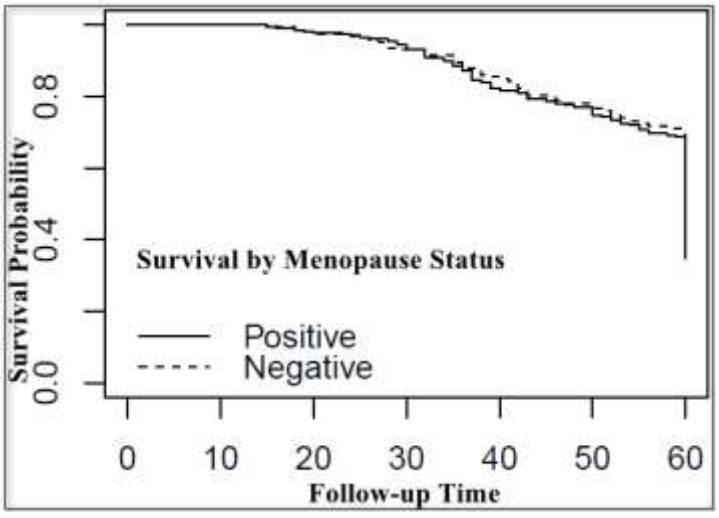

Figure 7. Survival Plot Stratified by Menopause after accounting for Competing Risk

Age is a key factor in the initiation of competing risk; thus, it is crucial to investigate the ideal age at which patients start to notice a substantial impact of competing risk events during breast cancer therapy. It is also important to highlight that

geographical differences in breast cancer are known to exist, necessitating a country-based approach to the age at which competing risks first appear. Figure 8 depicts the outcomes of optimization to determine the age at which competing risk first manifests itself. Figure 8 and Equation (5) make it clear that at the age of 57, breast cancer patients in Ghana start to notice a major influence of competing risks. This groundbreaking discovery closes a knowledge gap that will allow researchers employing product-limit estimators to restrict their study to people under the age of 58 for their analysis in Ghana. Additionally, this shows that competing risk models for Ghana's dataset are applicable at ages greater than 58 since, at these ages, competing risk events will start to be significant.

Table 4 shows the standard errors of KM estimates under the dataset for verified and unverified phenomena for the covariates recurrent, menopause, hospitalization, genetics, ER, PR, and HER2. The results reveal that, generally, the KM estimates are less biased with smaller standard errors when the dataset is verified for competing risk. The significance or otherwise of the p-values of both cases was however consistent with verified and unverified competing risk phenomena.The menopause status of BC patients was found to be insignificantly associated with survival. On the other hand, the status of BC patients with respect to recurrence, hospitalization, genetics, ER, PR, and HER2 were found to have significant p-values. The probabilities estimates of survival were overestimated for a dataset unverified of competing risk and found to have generally higher levels of standard error.

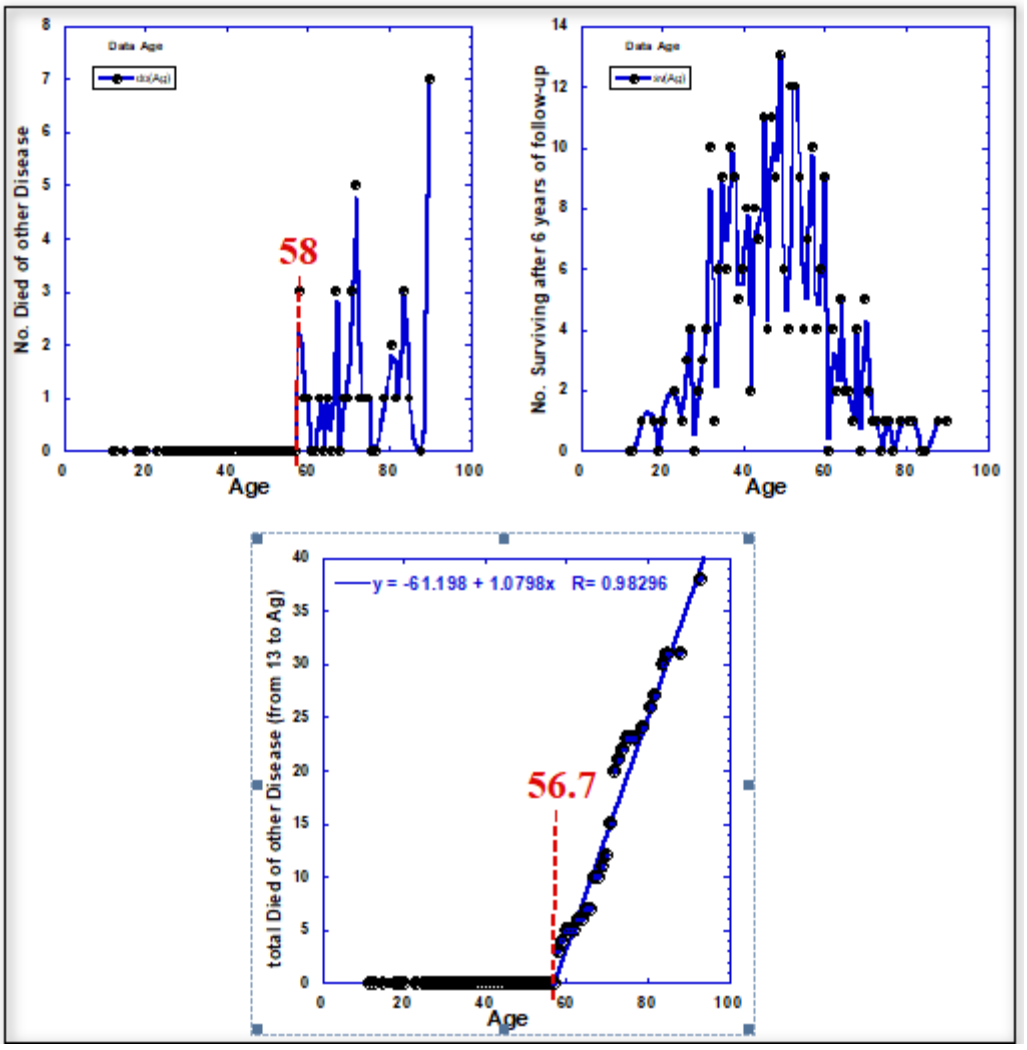

Figure 8. Plot to determine optimised aged for competing risk

We can estimate the total deaths of other disease Do,tot from 56.7 to the age (Ag), as follows:

$$D_{o,tot} = 1.0798 * (Ag - 56.7) \quad (5)$$

56.7 is the maximum theoretical age where the patient died with other diseases, which is null, in practice is $57 years =$

$58 - 1$.

The ROC curve has the ability to dichotomize into thresholds and compute the resulting sensitivity and specificity to forecast clinical risk by classifying patients into greater or lower risk categories of clinical importance. Figures 9 and 10, respectively, show the ROC curve analysis plots of verified or unverified datasets of competing risk for the patients' variables considered in the study. In general, the ROC analysis reveals a higher area under the ROC curve in favor of the dataset set verified for competing risks against that which did not. The standard errors were smaller for dataset verified for competing risk as opposed to those unverified. From the ROC analysis in Table 5, the covariates that drive breast cancer survivorship are, respectively, metastasis status (AUC=0.930), stage at diagnosis (AUC = 0.929), and tumor size (AUC = 0.929) and the number of lymph nodes (AUC=0.864) involved. Per the yardstick for explaining the ROC curve statistic, metastasis status, and stage at diagnosis, and tumor size were classified under outstanding discrimination (AUC > 0.9). The number of lymph nodes involved (AUC=0.864) was classified under excellent discrimination ($0.8 \geq$ AUC > 0.7). The hospitalization status (AUC = 0.659), grade at diagnosis (AUC = 0.658), and molecular subtype (AUC = 0.625) were classified under acceptable discrimination ($0.7 \geq$ AUC > 0.6).

Table 4. KM estimates under treated and untreated Dataset

Dataset Status		P-value (Wald Test)	se	exp(coef)	LCI	UCI
Treatment	Recurrent	<0.01	0.01	2.102	1.388	3.184
No Treatment		<0.01	0.02	2.478	1.75	3.508
Treatment	Menopause	0.6	0.016	1.075	0.8302	1.391
No Treatment		0.08	0.018	1.2516	0.9763	1.604
Treatment	Hospitalization	<0.01	0.013	3.319	2.441	4.513
No Treatment		<0.01	0.014	3.735	2.891	4.827
Treatment	Genetics	<0.01	0.016	0.3015	0.2318	0.3922
No Treatment		<0.01	0.017	0.3603	0.2828	0.459
Treatment	ER	<0.01	0.017	0.2887	0.2214	0.3765
No Treatment		<0.01	0.017	0.349	0.2737	0.445
Treatment	PR	<0.01	0.0016	0.2861	0.2191	0.3734
No Treatment		<0.01	0.0017	0.3426	0.2685	0.4373
Treatment	HER2	<0.01	0.015	1.651	1.259	2.163
No Treatment		<0.01	0.0016	1.493	1.105	2.018

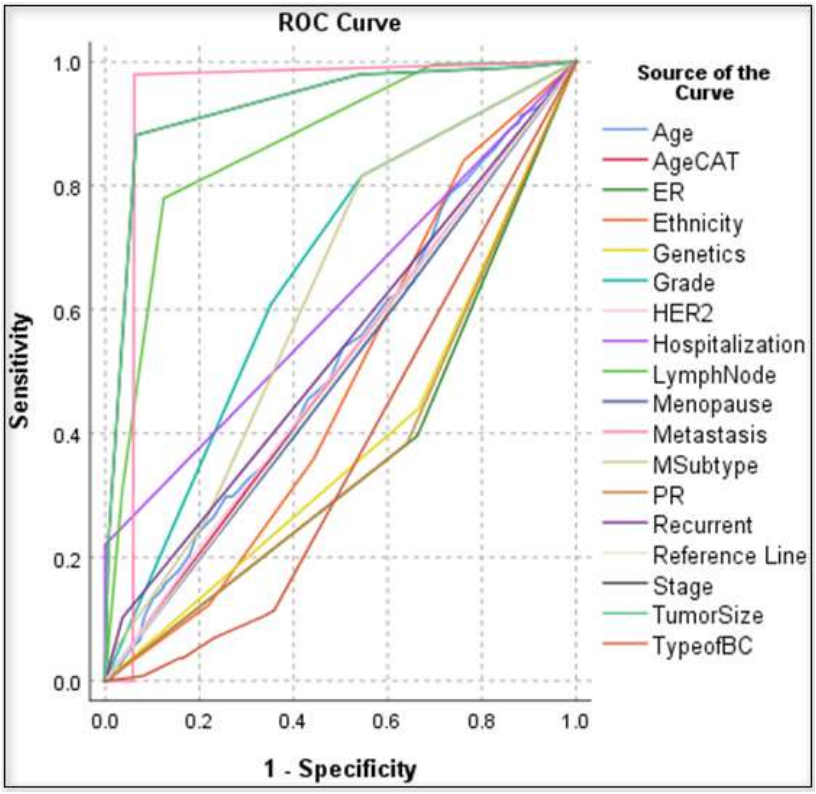

Figure 9. Plot of ROC curve Analysis for Unverified Competing Risk

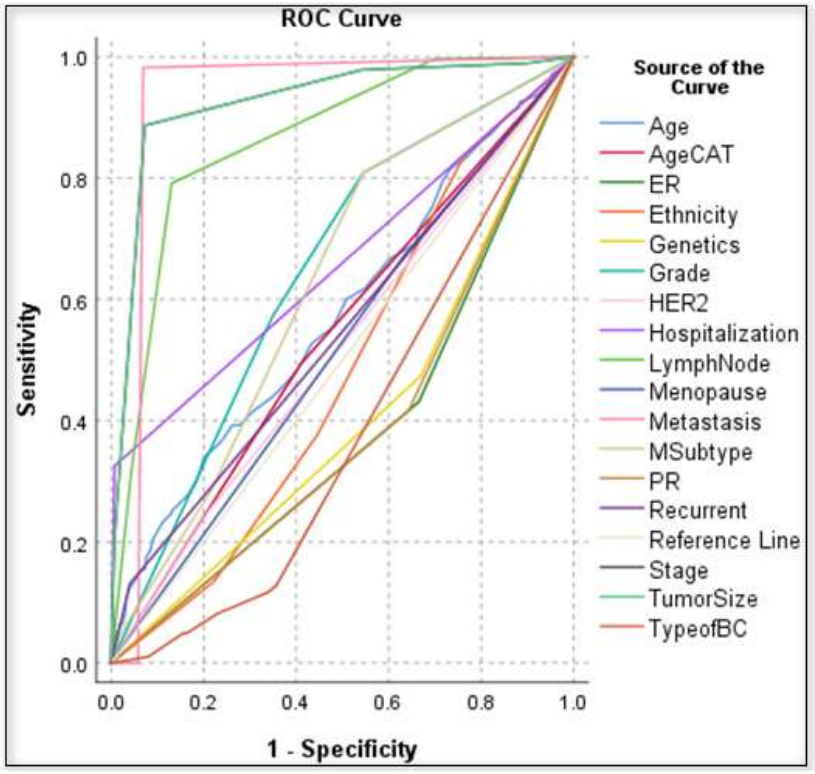

Figure 10. Plot of ROC curve Analysis for verified Competing Risk

Table 5. ROC Curve Analysis Output

Test Result Variable(s)	With Competing Risk Treatment					Without Competing Risk Treatment				
	Area under ROC curve	Std. Error	P-value	95% CI (LB)	95% CI (UB)	Area under ROC curve	Std. Error	P-value	95% CI (LB)	95% CI (UB)
Age	0.575	0.024	0.002	0.528	0.622	.520	.025	.431	.470	.570
AgeCAT	0.542	0.024	0.086	0.494	0.59	.505	.025	.848	.455	.555
Recurrent	0.545	0.024	0.062	0.498	0.593	.533	.025	.199	.483	.583
HER2	0.52	0.024	0.413	0.472	0.568	.508	.025	.758	.458	.558
ER	0.383	0.024	<0.01	0.336	0.429	.366	.025	.000	.318	.415
PR	0.387	0.024	<0.01	0.34	0.433	.371	.025	.000	.323	.420
MSubtype	0.625	0.024	<0.01	0.578	0.671	.622	.025	.000	.573	.670
Grade	.658	.024	.000	.611	.705	0.647	0.023	<0.01	0.601	0.693
TypeofBC	0.383	0.024	<0.01	0.336	0.43	.375	.024	.000	.327	.422
Stage	.929	.012	.000	.905	.953	0.926	0.012	<0.01	0.902	0.95
Metastasis	.930	.015	.000	.900	.959	0.928	0.015	<0.01	0.899	0.957
LymphNode	0.864	0.016	<0.01	0.833	0.895	.864	.016	.000	.833	.896
Menopause	0.519	0.024	0.429	0.471	0.567	.495	.025	.850	.445	.545
Ethnicity	0.482	0.025	0.468	0.434	0.53	.478	.025	.379	.428	.527
Hospitalization	0.659	0.023	<0.01	0.614	0.704	.610	.025	.000	.561	.659
TumorSize	.929	.012	.000	.905	.953	0.926	0.012	<0.01	0.902	0.95
Genetics	0.402	0.024	<0.01	0.355	0.449	.387	.025	.000	.338	.436

Figure 11 and Table 6 show the plot and estimations with respect to the survivorship of BC patients for each molecular subtype. From the plot, the triple negative molecular subtype was associated with the poorest prognosis, followed by Luminal B. The best prognosis was associated with patients who were linked with Luminal A. The survival plot looks at how long it takes for patients with the four-level molecular subtype to encounter the event (death from breast cancer). At periods 0–7 months into the research, 100% of the cohorts had not yet encountered the occurrence. There are four (4) different treatment groups (Luminal A, Luminal B, Triple Negative, and Human Epidermal Receptor 2), and each is denoted by a distinctively colored line. To calculate the survival time, five years of cohorts for each category were monitored. A fraction of the cohorts experiencing the event of interest are represented by the size of the vertical drop in the graph. Luminal A, the molecular subtype with the best prognosis, experienced fewer vertical declines over the whole follow-up period. The prognosis was worse for Triple Negative and Luminal B, which both `displayed significant vertical declines in the graph. The log rank (Mantel-Cox) approach compares the equality of survival functions by equally weighting each point in time. There were differences in molecular subtype levels that were statistically significant (chsq = 125.247; df = 3, p-value = 0.00). The triple negative molecular subtype has the lowest estimate (48.48), followed by the Luminal B and HER-2 molecular subtypes, both of which are below the overall estimate (52.773) and hence suggest a bad prognosis.

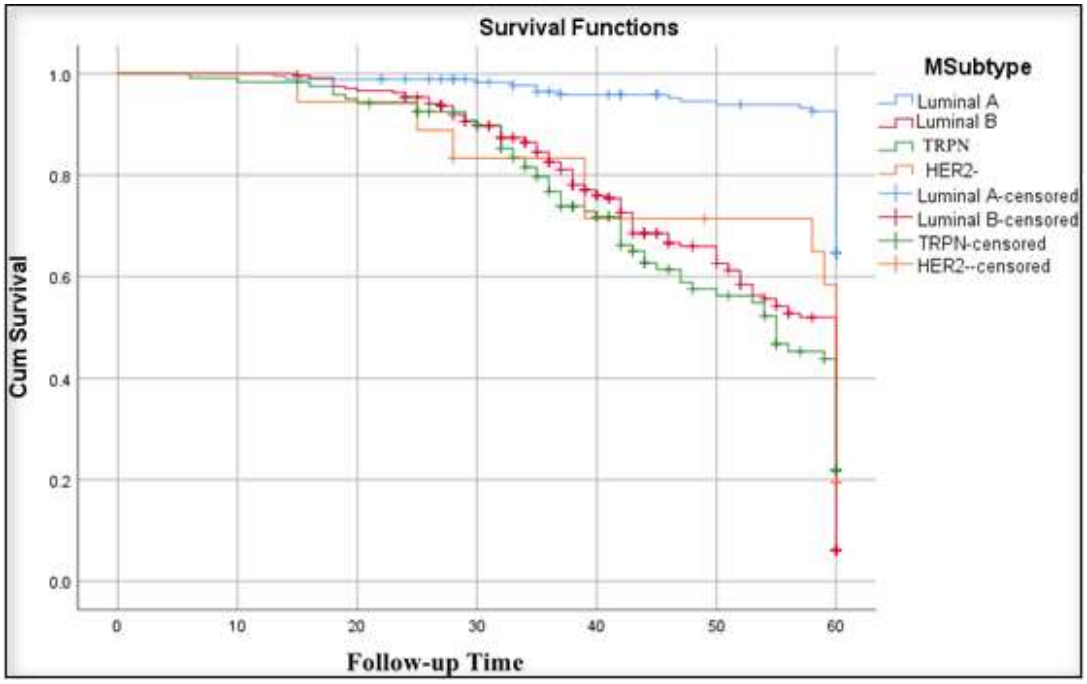

Figure 11. Survival Plot of Molecular Subtypes of Breast Cancer

Table 6. Survival Estimates for Molecular Subtype

Molecular Subtype	Estimate	Std. Error	Lower Bound (CI)	Upper Bound(CI)	Logrank Test	
Luminal A	58.42	0.519	57.402	59.438	statistic	5.247
Luminal B	50.349	0.867	48.649	52.049	P-value	<0.00
TRPN	48.48	1.326	45.881	51.08		
HER2-	51.083	3.588	44.05	58.116	df	3
Overall	52.773	0.536	51.722	53.823		

The survival plot and projections for BC patients for progesterone receptor status are shown in Figure 12 and Table 7. The plot implied that a progesterone receptor-negative state would have a worse prognosis. Patients with a positive status for the progesterone receptor had the best prognosis. The time it takes for patients with the two-level progesterone receptor status to experience the event is examined in the survival plot (death from breast cancer). At the beginning of the research, between 0 and 14 months, 100% of the cohorts had not yet experienced the event. Progesterone receptor positive or negative, there are two (2) distinct treatment groups, and each is shown by a line that is uniquely colored. Each group was followed for five years in order to determine the survival time. The size of the vertical drop in the graph represents a portion of the cohorts who experienced the relevant event. The best prognosis and fewer vertical drops in the plot were associated with progesterone receptor positivity over the entire follow-up period. There were statistically significant differences in progesterone status levels (chsq = 67.396; df = 1, p-value =0.00). The lowest estimate (50.143) for progesterone negative is below the overall estimate (52.773), indicating a poor prognosis.

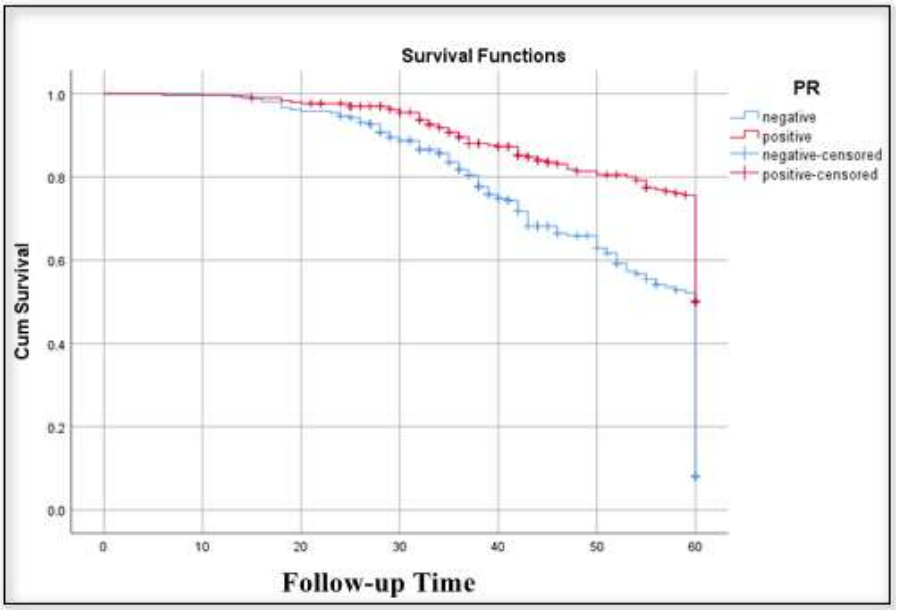

Figure 12. Survival Plot of Progesterone Receptor Status

Table 7. Survival Estimates for Progesterone Receptor

PR	Estimate	Std. Error	Lower Bound (CI)	Upper Bound(CI)	Logrank Test	
negative	50.143	.855	48.467	51.819	statistic	67.396
positive	54.972	.646	53.707	56.237	P-value	<0.00
Overall	52.773	.536	51.722	53.823	df	1

Figure 13 and Table 8 both display the survival plot and hospitalization status projections for BC patients. According to the plot, a BC patient's prognosis would be poorer if they were hospitalized. The best outcome was for breast cancer patients who were not hospitalized. The survival plot looks at how long it takes for patients with a two-level hospitalization status to encounter the event (death from breast cancer). 100% of the cohorts had not yet encountered the event at the start of the study, which took place between 0 and 8 months. The two (2) distinct treatment groups associated with hospitalization status are each represented by a line that is colored differently. Five years were spent monitoring each group to ascertain how long they survived. A part of the cohorts that experienced the relevant event is represented by the size of the vertical drop in the graph. Breast cancer patients with no hospitalizations over the whole follow-up period had the best prognosis and fewer vertical declines in the plot. Hospitalization status levels varied in a way that was statistically significant (chsq = 101.244; df = 1; p-value =0.00). A bad prognosis is indicated by the fact that the lowest estimate for hospitalized patients (45.629) is lower than the total estimate (52.773).

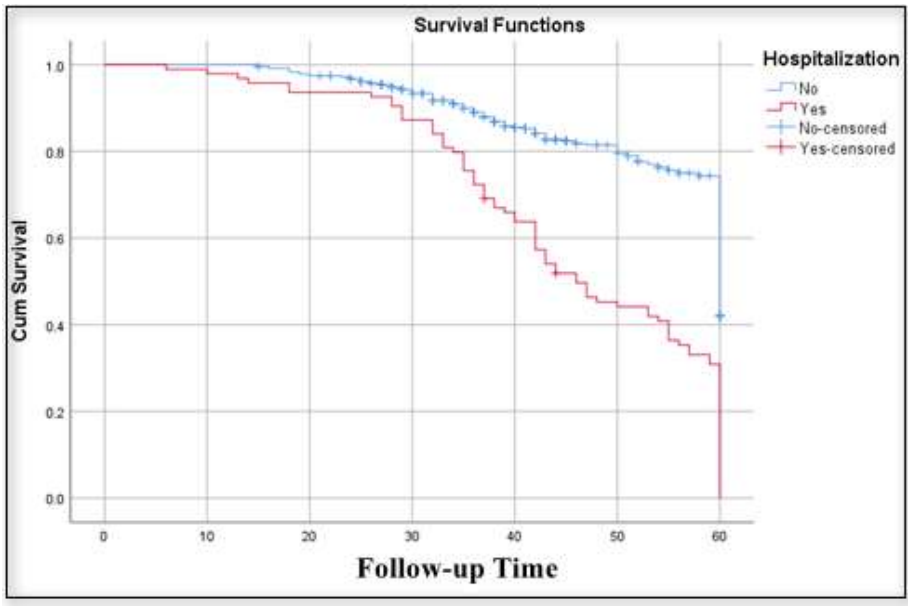

Figure 13. Survival Plot of Hospitalization Status

Table 8. Survival Estimates for Hospitalization

Hospitalization	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
No	54.489	.541	53.428	55.550	statistic	101.244
Yes	45.629	1.441	42.805	48.453	P-value	<0.00
Overall	52.773	.536	51.722	53.823	df	1

The survival plot and eostrogen-receptor status projections for BC patients are shown in Figure 14 and Table 9. A BC patient's prognosis would apparently be worse if they had an eostrogen-receptor negative status. Breast cancer patients with positive eostrogen-receptor status had the greatest results. At the commencement of the trial, which began between 0 and 7 months later, 100% of the cohorts had not yet experienced the occurrence. Each line that represents the two (2) unique treatment groups according to eostrogen-receptor status is colored differently. Each group was observed for five years to determine how long they persisted. The size of the vertical drop in the graph represents a portion of the cohorts that were exposed to the primary event. The breast cancer patients with the best prognosis and the fewest vertical decreases in the plot were those with eostrogen-receptor positive status during the whole follow-up time. There were statistically significant variations in eostrogen receptor levels (chsq = 64.618; df =1; p-value = 0.00). The fact that the lowest estimate for eostrogen-receptor-negative patients (50.330) is lower than the overall estimate suggests a poor outlook (52.773).

Table 9. Survival Estimates for Eostrogen Receptor Status

ER	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
negative	50.330	.851	48.663	51.997	statistic	64.618
positive	54.696	.663	53.396	55.996	P-value	<0.00
Overall	52.773	.536	51.722	53.823	df	1

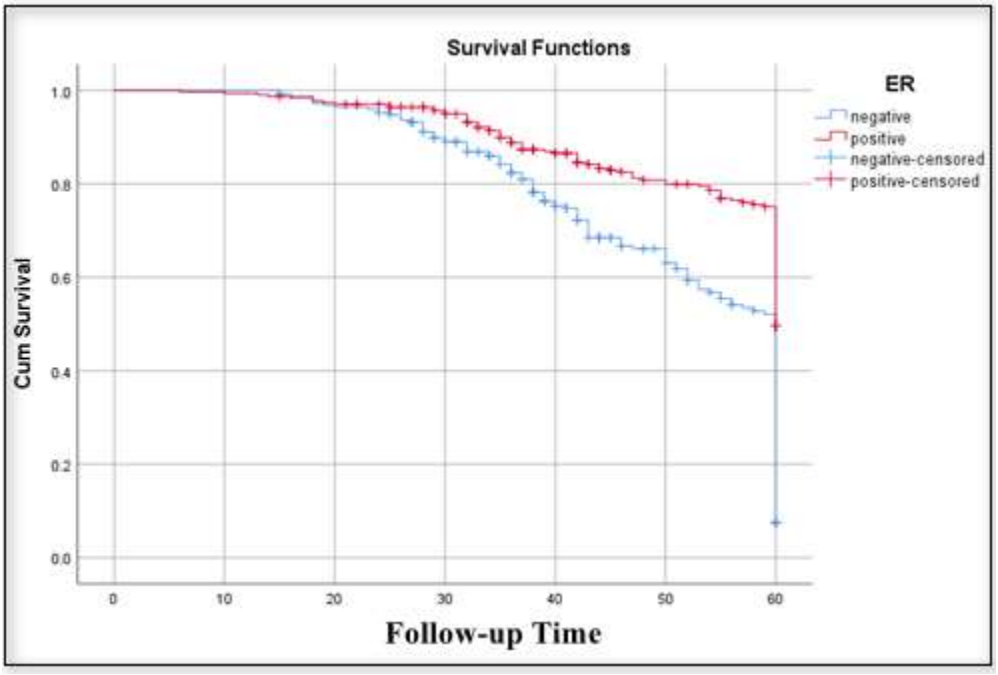

Figure 14. Survival Plot of Eostrogen Receptor Status

Figure 15 and Table 10 display the survival plot and Human Epidermal Receptor-2 (HER2) status projections for BC patients. If a patient has an HER-2-positive status, their prognosis is likely to be worse. Patients with HER-2-positive breast cancer had the greatest outcome. 100% of the cohorts had not yet experienced the event from the beginning of the study through the sixth month. According to HER-2 status, each line representing the two (2) distinct treatment groups is colored differently. Five years were spent observing each group to see how long they persisted. A fraction of the cohorts exposed to the main event is represented by the size of the vertical drop in the graph. Patients with HER-2 -negative status throughout the whole follow-up period had the best prognosis and the fewest vertical reductions in the plot. HER-2 levels varied statistically significantly (chsq =15.988; df =1; p-value = 0.00). Poor prognosis is suggested by the fact that the lowest estimate for HER-2- positive patients (48.964) is lower than the overall estimate (52.773).

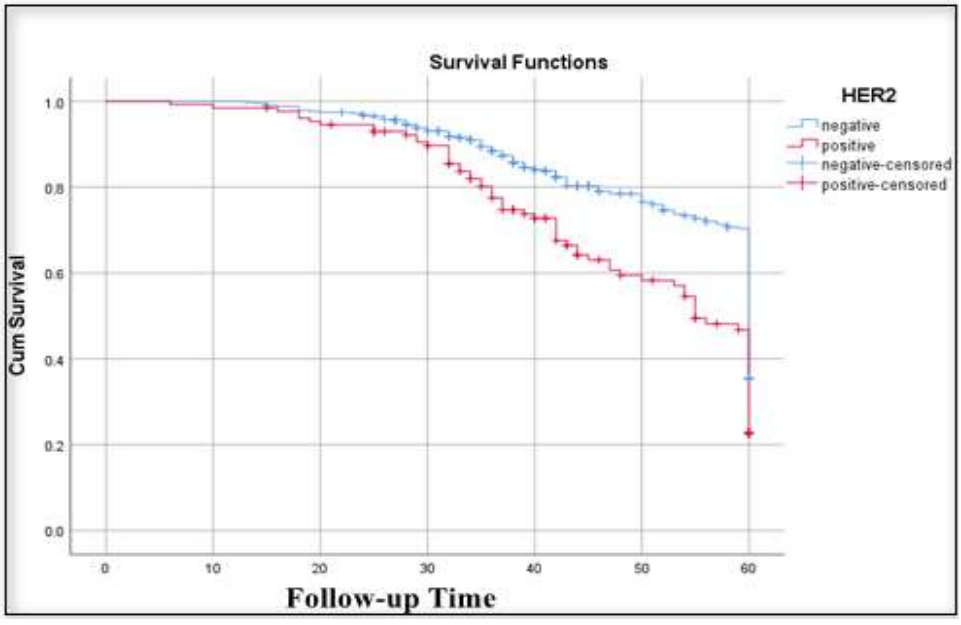

Figure 15. Survival Plot of Human Epidermal Receptor2 Receptor Status

Table 10. Survival Estimates for Human Epidermal Receptor2 Status

HER2	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
negative	53.867	.573	52.744	54.989	statistic	15.988
positive	48.964	1.278	46.459	51.470	P-value	<0.00
Overall	52.773	.536	51.722	53.823	df	1

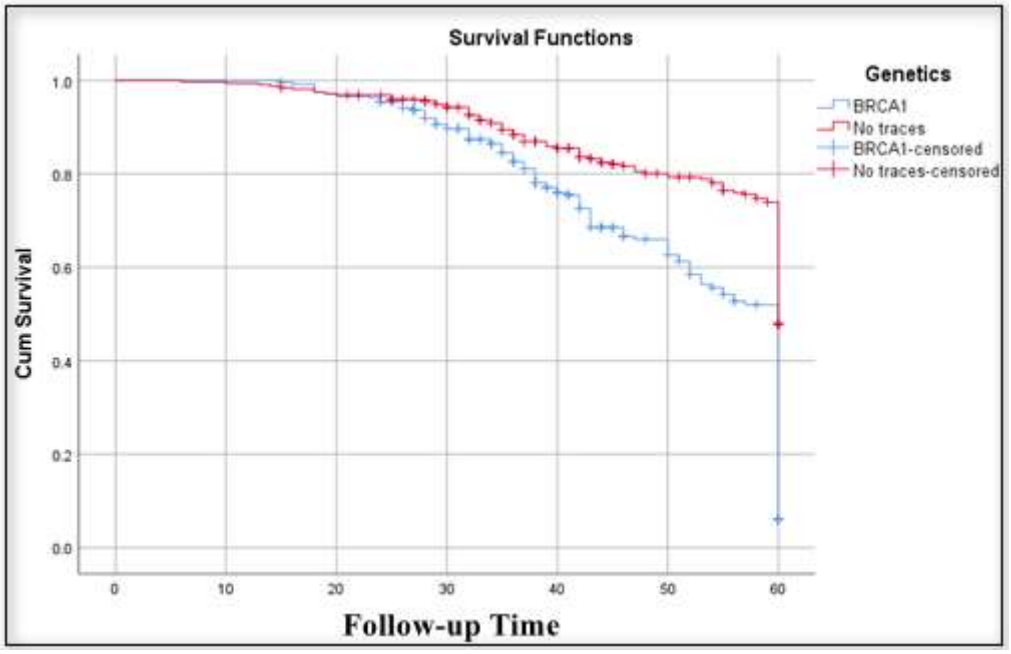

Figure 16. Survival Plot of Genetic Status

The survival plot and genetic status projections for BC patients are shown in Figure 16 and Table 11. A patient's prognosis is likely to be worse if they have genetic (BRCA 1) breast cancer. The best results were obtained by patients whose breast cancer was not genetically connected. From the start of the trial until the 10th month, the event had not yet occurred in 100% of the cohorts. Each line denoting the two (2) distinct treatment groups is colored differently depending on genetic status. Each group was watched for five years to see how long they lasted. A fraction of the cohorts exposed to the main event is represented by the size of the vertical drop in the graph. Patients without genetically related BC over the entire follow-up period had the best prognosis and the fewest vertical drops in the plot. Statistics showed a statistically significant variation in genetic status levels (chsq = 59.429; df =1; p-value = 0.00). The fact that the lowest estimate for genetically related BC patients (50.349) is lower than the overall estimate points to a poor outlook (52.773).

Table 11. Survival Estimates for Genetic Status

Genetics	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
BRCA1	50.349	.867	48.649	52.049	statistic	59.429
No traces	54.432	.664	53.130	55.734	P-value	<0.00
Overall	52.773	.536	51.722	53.823	df	1

For patients with BC, Figure 17 and Table 12 display the survival plot and tumor size projections. Breast cancer survivorship with respect to tumor sizes between >60mm and >50mm<60mm dropped below the overall survival time of 52.773, which indicates that a patient's prognosis is likely to be poorer at higher tumor sizes.

The event was postponed in 100% of the cohorts from the beginning of the experiment to the seventh month. Depending on the size of the tumor, each of the five (5) separate treatment groups' lines is colored differently. Five years of observation were used to gauge each group's longevity. The size of the vertical drop in the graph represents a portion of

the cohorts exposed to the primary event. Since their survival estimates were higher than the average estimates of 52.773, patients with tumor diameters between >5mm $\leq$ 10mm, >10mm $\leq$ 20mm, and >20mm $\leq$ 50mm for the full follow-up period had the best prognosis and the fewest vertical declines in the plot. According to the estimations in Table 10, survival rates tend to decline as tumor size increases. Tumor size levels varied statistically significantly (chsq = 265.344; df = 4 ; p-value = 0.00), according to statistics.

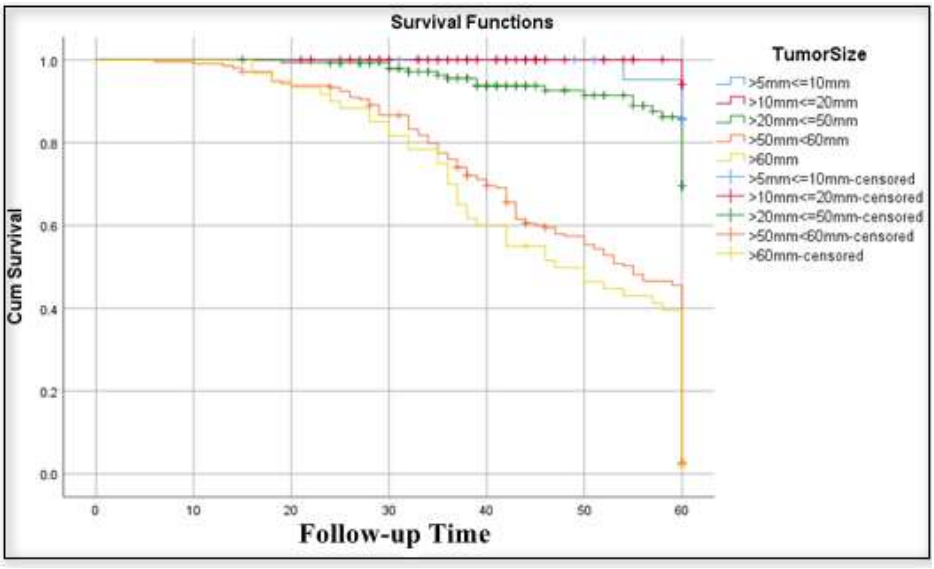

Figure 17. Survival Plot of Tumor Size

Table 12. Survival Estimates for Tumor Size

TumorSize	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
>5mm $\leq$ 10mm	59.714	.341	59.045	60.384	statistic	265.344
>10mm $\leq$ 20mm	60.000	.000	60.000	60.000	P-value	<0.00
>20mm $\leq$ 50mm	57.845	.623	56.625	59.066		
>50mm<60mm	47.951	.981	46.028	49.875		
>60mm	45.825	1.869	42.162	49.488	df	4
Overall	52.773	.536	51.722	53.823		

Figure 18 and Table 13 show the survival plot and an estimation of the involvement of lymph nodes for BC patients. Breast cancer survivorship with regard to lymph node involvement fell below the overall survival time of 52.773 when cancer has spread to four to nine underarm lymph nodes,(level 3); cancer has spread to ten or more axillary lymph nodes, and one site is larger than two millimeters (level 2), which indicates that a patient's prognosis is poorer. The prognosis was often better for patients whose cancer had progressed to one to three underarm lymph nodes (level 1). From the start of the experiment through the seventh month, 100% of the cohorts had the event postponed. Each of the three (3) distinct treatment groups' lines is colored differently depending on the degree of lymph node involvement. Each group's longevity was evaluated over a period of five years of monitoring. The estimates in Table 11 show that as the number of lymph nodes involved rises, survival rates tend to decrease. Statistics revealed statistically significant variation in lymph node involvement levels (chsq = 183.929; df = 3, p-value = 0.00).

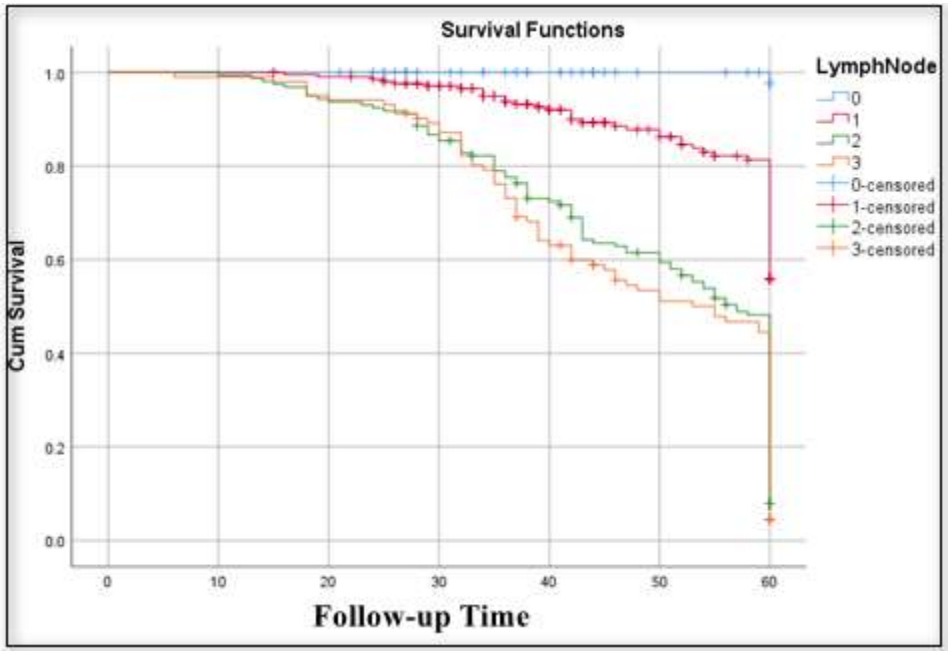

Figure 18. Survival Estimates for Lymph Node Involvement

Table 13. Survival Estimates for Tumor Size

Lymphnode	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
1	56.574	.640	55.320	57.828	P-value	<0.00
2	48.713	1.132	46.495	50.931	statistic	183.929
3	47.331	1.421	44.546	50.116	df	3
Overall	52.773	.536	51.722	53.823		

The survival plot and an estimate of the state at diagnosis for BC patients are shown in Figure 19 and Table 14. Breast cancer survivorship based on stage at diagnosis went below the overall survival time of 52.773 for the higher stages of stages III and IV, with estimations of 47.951 and 45.825, respectively, which shows that a patient's prognosis is worse. Patients with cancer in the earlier stages of 0, I, and II often had a better outlook since their survival estimates were higher than the overall survival estimate. According to the projections in Table 12, survival rates tend to decline as BC staging increases. Statistics showed that there were statistically significant differences in the amounts of lymph node involvement (chsq = 265.344; df = 4, p-value = 0.00).

Table 14. Survival Estimates for State at Diagnosis

Stage	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
0	59.714	.341	59.045	60.384	P-value	<0.00
1	60.000	.000	60.000	60.000	Statistic	265.344
2	57.845	.623	56.625	59.066		
3	47.951	.981	46.028	49.875	Df	4
4	45.825	1.869	42.162	49.488		
Overall	52.773	.536	51.722	53.823		

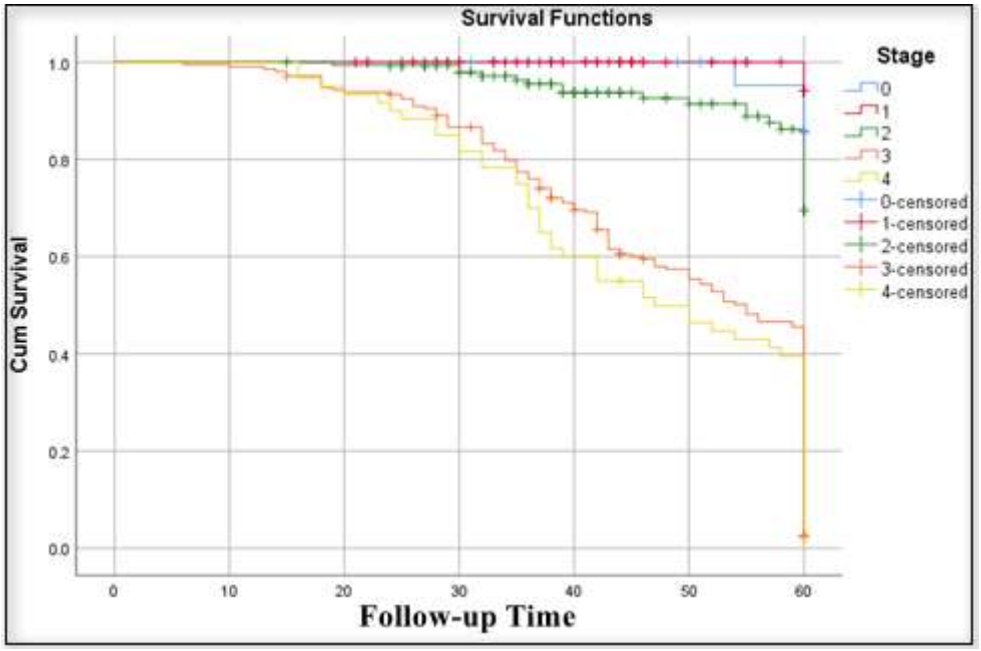

Figure 19. Survival Plots for State at Diagnosis

The survival plot and grade at diagnosis forecasts for BC patients are shown in Figure 20 and Table 14. Survival tends to drop as the grade upon diagnosis rises. However, the calculations showed that patients with Grade II status had the worst prognosis, while those with grade I status had the highest survival rates. From the start of the trial until the seventh month, 100% of the cohorts had not yet encountered the occurrence. There were statistically significant differences in grade levels (chsq = 265.344; df = 2; p-value = 0.00).

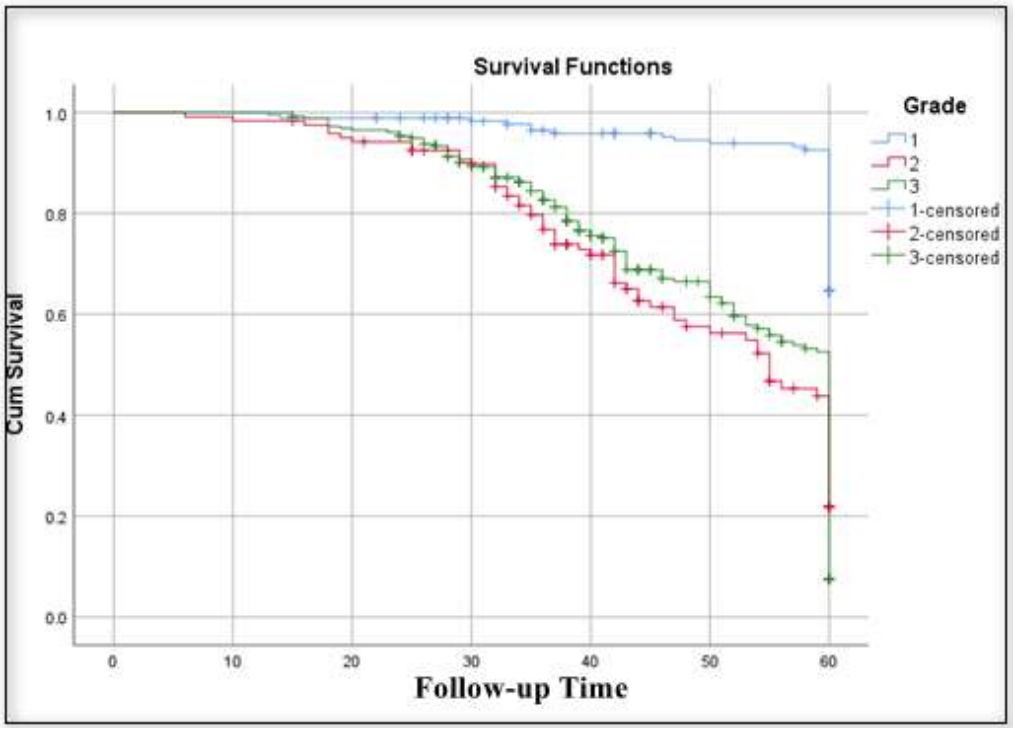

Figure 20. Survival Plots for Grade at Diagnosis

Table 14. Survival Estimates for Grade at Diagnosis

Grade	Estimate	Std. Error	Lower Bound (CI)	Upper Bound (CI)	Logrank Test	
1	58.420	.519	57.402	59.438	statistic	265.344
2	48.480	1.326	45.881	51.080		
3	50.443	.842	48.793	52.093	df	2
Overall	52.773	.536	51.722	53.823	P-value	0.00

4. Conclusion

In a plethora of recent studies, single-event models have been applied without any attempt to verify the underlying hypotheses that support their robustness. This error in the literature causes bias to increase and compromises the reliability of the anticipated estimations. The study's goal was to demonstrate the bias that results from analyzing a particular event of interest using a cause-specific product-limit estimate while disregarding competing risks. To distinguish between ages with and without competing events, competing risk models were contrasted with the complement of the Kaplan-Meier estimator. Where the complements of the Kaplan-Meier estimates were higher than the estimates of the cumulative incidence functions, a bias of overestimation was thought to exist. The premise that the product-limit estimator can generate accurate estimates would be satisfied by weaning off these competing events. When the estimates from the two datasets were compared, it became clear from ROC curve analysis and standard error analysis that the estimates from the survival dataset, which has been verified for competing risk, have smaller standard errors and a larger area under the ROC curve, increasing their robustness and power. By doing this, a proper foundation for the estimates reached has been established, making them more trustworthy than the estimates of survivorship in numerous studies that did not take into account the underlying assumptions of the product-limit estimator. It is crucial to specify the nature of the survival dataset, make it a requirement for analysis, and determine whether it contains competing events or not. This will provide justification and guidance for the statistical approach to be used and prevent a common literature error. The ideal age for Ghanaian BC patients to start developing comorbidities has been identified as 58 years. Future researchers will be able to use non-competing risk models like the Kaplan-Meier and Cox PH Model in a defined non-competing risk domain (57 years and under) as well as analyze competing risk events in a defined set (58 years and above).

References

Abdel-Qadir, H., Fang, J., Lee, D. S., Tu, J. V., Amir, E., Austin, P. C., & Anderson, G. M. (2018). Importance of Considering Competing Risks in Time-to-Event Analyses: Application to Stroke Risk in a Retrospective Cohort Study of Elderly Patients with Atrial Fibrillation. *Circulation Cardiovascular Quality and Outcomes Journal, 11*(7), 87-109. https://doi.org/10.1161/CIRCOUTCOMES.118.004580

Adeloye, D., Sowunmi, O. Y., Jacobs, W., David, R. A., Adeosun, A., & Amuta, A. O. (2018). Estimating The Incidence of Breast Cancer in Africa: A Systematic Review and Meta-Analysis. *Journal of Global Health, 8*(1), 1-21. https://doi.org/10.7189/jogh.08.010419

Arani, M. D., Abadi, A., Yavari, A., Bashiri, Y., Liley, M., & Bajdik, C. (2018). Evaluation of risk factors in patients with breast cancer III and IV: Comparison of Cox and Fine-Grey competing risk regression models. *Biomedical Research and Therapy Journal, 5*(2), 2022-2033. https://doi.org/10.15419/bmrat.v5i02.417

Austin, P. C., & Fine, J. P. (2017). Practical recommendations for Reporting Fine-Gray Model Analyses for Competing Risk Data. *Statistics in Medicine Journal, 36,* 4391-4400. https://doi.org/10.1002/sim.7501

Bosson-Amedenu, S., Acquah, J., Nyarko, C. C., & Ouerfelli, N., (2022). A Comparison between Accelerated Failure Time Models in Analyzing the Survival of Breast Cancer Patients. *Journal of Cancer and Tumor International, 12*(1), 16-28. https://doi.org/10.9734/jcti/2022/v12i130166

Boyle, P. (2012). Triple-negative Breast Cancer: Epidemiological considerations and Recommendations. *Annals of Oncology, 23*(6), 24-33. https://doi.org/10.1093/annonc/mds187

Chapman, J. A., Pritchard, K. I., Goss, P. E., Ingle, J. N., Muss, H. B., Dent, S. F., … Gelmon, K. A. (2014). Competing risks of death in younger and older postmenopausal Breast Cancer Patients. World Journal of Clinical Oncology, 5(5), 1088 – 1096. https://doi.org/10.5306/wjco.v5.i5.1088

Collet, D. (2015). Modelling Survival Data in Medical Research. Third Edition, Taylor and Francis Group Press, Florida, 2742. https://doi.org/10.1201/b18041

Ghavami, V., Mahmoudi, M., Foroushani, A. R., Baghishani, H., Shandiz, F. H., & Yaseri, M. (2017). Long-Term

Disease-free Survival of Non-Metastatic Breast Cancer Patients in Iran: A survival Model with Competing Risks Taking Cure Fraction and Frailty into Account. Asian Pacific Journal of Cancer Prevention, 18.

Glas, N. A., Mandy, K., & Vandenbroucke, J. P. (2016). Performing Survival Analyses in the presence of Competing Risks: A clinical Example in Older Breast Cancer patients. *Journal of the National Cancer Institute, 108*(4). https://doi.org/10.1093/jnci/djv366

Kleinbaum, D. G, & Klein, M. (2015). Survival Analysis: A Self-Learning Text. Second Edition, Springer Science Publishers, United States of America, 1020.

Koller, M. T., Raatz, H., Steyerberg, E. W., Wolbers, M. (2011). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine, 31*(11), 1089–1097. https://doi.org/10.1002/sim.4384

Mensah, Y. B. (2021). A Review of Computed Tomography Patterns of Metastatic Breast Cancer Patients Undergoing Treatment at a Private Oncology Centre in Ghana. *Pan African Medical Journal, 38*(50), 101-110. https://doi.org/10.11604/pamj.2021.38.50.21945

Ohene-Yeboah, M., & Adjei, E. (2012). Breast Cancer in Kumasi, Ghana. *Ghana Medical Journal, 46*(1), 8–13.

Okifo, F. O., & Tuoyire, A. D. (2021). Breast Cancer Treatment and Outcomes at Cape Coast Teaching Hospital, Ghana. *Ghana Medical Journal, 55*(3), 190-197. https://doi.org/10.4314/gmj.v55i3.3

Okifo, F. O., & Tuoyire, A. D. (2021). Breast Cancer Treatment and Outcomes at Cape Coast Teaching Hospital, Ghana. *Ghana Medical Journal, 55*(3), 190-197. https://doi.org/10.4314/gmj.v55i3.3

Osei-Afriyie, S., Addae, A. K., Oppong, S., Amu, H., Ampofo, E., & Osei, E. (2021). Breast Cancer Awareness, Risk Factors and Screening Practices among Future Health Professionals in Ghana: A Cross-Sectional Study. *PLoS ONE, 16*(6), 23-42. https://doi.org/10.1371/journal.pone.0253373

Perera, D., & MUditha, V. (2017). *Statistical Analysis and Modeling of Ovarian and Breast Cancer.* University of South Florida Journal, Graduate Thesis, 143.

Pham, M. H. & Kafle, R. C. (2016). Competing Risks Analysis of African American Breast Cancer Patients. *Advances in Breast Cancer Research, 6*(1), 28–41. https://doi.org/10.4236/abcr.2017.61003

Quayson, S. E., & Wiredu E. K (2014). Breast Cancer in Accra, Ghana. *Journal of Medical and Biomedical Sciences, 3*(3), 21-26. https://doi.org/10.4314/jmbs.v3i3.4

Sapir-Pichhadze, R., Pintilie, M., & Tinckam, K. J. (2016). Survival Analysis in the Presence of Competing Risks: The Example of Waitlisted Kidney Transplant Candidates. *American Journal of transplantation, 16*(7), 1958–1966. https://doi.org/10.1111/ajt.13717

Seshie, B., Adu-Aryee, N. A., & Dedey, F. (2016). Retrospective Analysis of Breast Cancer Subtype Based on ER/PR and HER2 Status in Ghanaian Patients at the Korle Bu Teaching Hospital, Ghana. *BMC Clinical Pathology, 15*(14). https://doi.org/10.1186/s12907-015-0014-4

Stark, A., Kleer, C. G., & Martin, I. (2016). African Ancestry and Higher Prevalence of Triple-Negative Breast Cancer: Findings from an International Study. Cancer, (116), 4926- 4932. https://doi.org/10.1002/cncr.25276

Swapna, D. (2013). *Computing the Competing Risks' Modeling Survival Data with Competing Risk Events using SAS Macros.* Cytel Publishers, Sweden, 645.

Thomas, A. S., Kidwell, M. K., & Oppong, J. K. (2017). Breast Cancer in Ghana: Demonstrating the Need for Population-Based Cancer Registries in Low- and Middle-Income Countries. *Journal of Global Oncology, 3*(6). https://doi.org/10.1200/JGO.2016.006098

Usman, M., Dikko, H. G., Bala, S., & Gulumbe, S. U. (2014). An Application of Kaplan–Meier Survival Analysis Using Breast Cancer Data. Sub-Saharan African. *Journal of Medicine, 1*(3), 132-141. https://doi.org/10.4103/2384-5147.138940

Wright, C. M., Nowak, A. K., & Halkett, G. (2020). Incorporating Competing Risk Theory into Evaluations of Changes in Cancer Survival: Making the Most of Cause of Death and Routinely Linked Socio-demographic Data. *BMC Public Health, 20*(1), 10- 22. https://doi.org/10.1186/s12889-020-09084-8

Copyrights