Review Journal of Social Psychology & Social Works
http://socialworksreview.com
ISSN-E: 3006-4724 Volume: 3
ISSN-P: 3006-4716 Issue: 1 (2025)

A Comparative Study among Parametric, Semiparametric and Non-parametric Techniques Using Survival Data

Inam Ur Rahman[1*], Nasir Ali[2], Abid Hussain[3], Mehvish Raja[4]

1. Department of statistics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan
*Email: inamurrahman85@gmail.com
2. Department of statistics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan
Email: nasir_stat@uaar.edu.pk
3 Department of statistics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan
Email: abid0100@gmail.com
4. Department of Economics, Quaid-e-Azam University, Rawalpindi, Pakistan
Email: mehvish.raja2080@gmail.com

DOI: https://10.71145/rjsp.v3i1.164

Abstract
This study offers a comparative analysis of parametric, semi-parametric, and non-parametric techniques of survival data analysis. It focuses on assessing the methods in terms of their efficiency in estimating survival probabilities and hazard functions. Alternate real and simulated datasets are then utilized to assess the relative strengths and weaknesses of different approaches with respect to efficiency, flexibility, and interpretability. The results indicate that the choice of technique is a factor of the underlying data characteristics, with parametric models working quite well under specific assumptions, semi parametric methods balancing between the structure and flexibility, while the nonparametric ones have the best performance in scenarios driven by the data. The AIC and BIC values determined from the model selection process demonstrated that the model finding fits the best under the given conditions as it has the smallest AIC and BIC scores. This is because AIC and BIC balance between goodness of fit and complexity in the model. For the Cox proportional hazard model results, most variables tested in the analysis showed very few statistically significant p- values. Results from Kaplan-Meier survival analysis strengthened the model by providing critical survival probability key stages over time, indicating the most occurrences beyond 60 months marked significant declines in survival rates. Hence, since the determined model com- bines predictive accuracy with interpret-ability, it should have as few but robust elements as would be meaningful in survival prediction without over-fitting. Therefore, the final model selection is based on the best compromise between explanatory power and statistical validity.

Keywords: Breast Cancer Survival, Kaplan-Meier, Cox Model, Parametric Models, Weibull, Exponential, Lognormal, Log Logistic

Introduction

Breast cancer is the most commonly diagnosed cancer among women and a leading cause of cancer-related death globally Ferlay et al. (2021). The survival rates, considered together, vary widely due to several clinical and pathological characteristics: tumor size, lymph node involvement, hormonal receptor status, and efficacy of treatment regimens Siegel et al. (2023). Understanding these factors is crucial for developing personalized treatment designs and, thus, improving patient outcomes. Survival analysis provides a statistical method through which one may consider the time-to-event data in medical research, and such consideration is critical in predicting the survival of patients for decision-making Collett (2015).

Survival analysis methods are broadly categorized into non-parametric, semi-parametric, and parametric models, with the **Kaplan-Meier estimator** being considered the most popular non-parametric method for estimating the survival probabilities over time without assuming an underlying probability distribution Kaplan and Meier (1958). In this case, survival functions are commonly depicted or compared between groups in clinical studies, while they ignore the effects posed by covariates. The limitation in this regard is solved by the use of the *Cox proportional hazards model, * which is semi-parametric Cox (1972). This model for instance provides an estimate of hazard ratios for the effect of covariates, while using as few assumptions as possible about the baseline hazard function and thus remains one of the most widely-used survival analysis techniques within epidemiologic studies Therneau and Grambsch (2000).
On the one hand, Cox models have the advantage of being flexible; however, they do not assume a particular survival distribution. Their prediction may therefore lack accuracy in some datasets. On the other hand, parametric survival models like the **Weibull, log-normal, log-logistic and exponential distributions** lend themselves to a more formal approach resting on the assumption of a known probability distribution for survival times Lawless (2011). Therefore, they are often more suitable in applications when the true survival function aligns with a specific distribution. On the other hand, the **Weibull model** dominates the two domains of reliability and medical research were increasing and decreasing hazard rates need to be modeled Meeker and Escobar (1998). The **log-normal model** applies well in situations where survival times are right-skewed, and the **log-logistic model** is appropriate when hazard rates increase initially and decrease later Klein and Moesch Berger (2003).
Kaplan-Meier estimation, Cox regression, and several parametric models (Weibull, log-normal, log- logistic, and exponential) have been applied for breast cancer survival data. The main objective is to compare these techniques as well as to assess their effectiveness in predicting survival probabilities and determining factors that significantly affect breast cancer mortality. Models chosen based on statistical criteria such as **Akaike Information Criterion (AIC)**, **Bayesian Information Criterion (BIC)**, etc. help to find the best-fitting model for survival analysis Burnham and Anderson (2002).
By comparing different techniques of survival analysis, a research goal will definitely gain insight into breast cancer in the future, as far as scrutinizing the statistical methodologies used in medical research is concerned. Results in this study will go a long way in equipping clinicians and researchers towards picking the most appropriate patient survival prediction models and, thus, making more informed treatment decisions.
There are many problems that arise in analyzing data. Broadly, these are as follows: (1) How to select a statistical model for estimation of parameters and (2) how to relate biological

information with statistical modeling. To tackle these problems, efficient methods and computational software have their utility. We can state a mathematical expression in terms of probability as: to analyze the survival time of a patient till time point t.

$S(t) = P[T > t]$. (1)

To cope with these challenges, you need methods and computer packages that work efficiently. In free R, package" survival" is used to apply these methods. The survival package provides functions for a CPH model, KM method, and parametric models for discrete and continuous outcomes.

In this paper we analysis the effect of breast cancer by using different variable. The section 1 represent the importance of survival analysis in the field of medical breast cancer. In this section 2 we talk about method and formulas here we use. The third section 3 is about data and evaluation criteria. The fourth section 4 is about Results and discussion. The fifth section 5 about the conclusion of this research paper.

Method and Technique

Kaplan-Meier Estimation Method

In breast cancer survival analysis, the Kaplan-Meier estimate is generally upheld as a method in which survival probabilities are estimated for various treatment regimens and to allow for the handling of censored data-cases in which the patient is lost to follow-up or is alive at the end of the study Kaplan and Meier (1958). Treatment would then be evaluated by estimating differences in survival among patients receiving different treatments (e.g., chemotherapy versus surgery), thereby calculating the effect of factors like tumor stage, hormone receptor status, and lymph node involvement on survival Blanche et al. (2013). The Kaplan-Meier survival function is expressed as:

$$S(i) = \prod_{g:.g.j \le i} \left(1 - \frac{K_h}{N_g}\right) \qquad (2)$$

Where i_g is the observed time for the g-th event, K_h is the number of events (e.g., deaths) that occurred at time i_g, and X_i is the number of individuals at risk just before time.

The Kaplan-Meier survival curve is stepwise, declining at each event time, with censored observations marked on the plot. The log-rank test assesses whether survival distributions differ significantly between groups Peto et al. (1977).

Cox Proportional Hazards model

In breast cancer studies, the Cox proportional hazard model analyzes the relationship between survival times and one or more variables like age, tumor size, hormone receptors, and lymph node involvement. This semi-parametric model is advantageous since it makes no assumptions concerning the nature of the underlying baseline hazard function and thus can concentrate on explaining the effects of the explanatory variables and also accepts censored data Cox (1972). Within the model, it is assumed that a hazard ratio for two individuals remains constant with respect to time and is defined as:

$$k(i|X) = k_0(i).\, e^{\beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 \ldots\ldots\ldots \beta_n y_n} \qquad (3)$$

Where $h(j|\boldsymbol{Y})$: Hazard rate at time j for an individual with covariates $\boldsymbol{Y}, h0(j)$: Baseline hazard function meaning hazard when all covariates are zero, $\beta 1, \beta 2, \ldots, \beta n$ are regression coefficients defining the effect of covariates.
In breast cancer, the Cox model permits clinicians to evaluate the influence of many risk factors concurrently, thus facilitating the identification of patients at high risk of recurrence or death Bland & Altman (1998). The model is strongly applied in any clinical research and leads to more individualized therapy for breast cancer patients.

Parametric Models

The Cox Proportional Hazards (CPH) model assesses the influence of the predictors on the time deviation from some event. It assumes that the effect of these predictors is time-constant-proportional. Parametric survival models may also be ill-suited if the random variable's distribution is normal. However, some are better suited to be chosen from among an array of parametric distributions: the Weibull, exponential, log-normal, and log-logistic distributions. The exposure to the Weibull distribution can thus be stated as:

$$f(y|\sigma,p) = \frac{p}{\sigma}\left(\frac{y}{\sigma}\right)^{p-1} e^{\left(-\frac{y^p}{\sigma}\right)}, \quad y > 0, \sigma > 0, p > 0 \qquad (4)$$

The exponential distribution is expressed as

$$f(x|\sigma) = \frac{1}{\sigma} e^{\left(-\frac{x}{\sigma}\right)}, \quad x > 0, \sigma > 0 \qquad (5)$$

The log-normal distribution is defined as

$$f(x|\mu,\sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)}, \quad x \in (0,\infty) \qquad (6)$$

the log-logistic distribution is given by

$$f(x|\alpha,\beta) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^2}, \quad x > 0, \alpha > 0, \beta > 0 \qquad (7)$$

Limitations of Clinical Data for Kaplan-Meier, Cox Hazard, and Parametric Models

Survival analysis is an integral part of clinical research, and several models such as the Cox proportional hazards model, the Kaplan–Meier estimator, parametric survival models, and Bayesian approaches provide a basis within which clinical investigators can assess survival time data. Survival analysis methods, however, come with their own distinct limitations. The Cox model Cox (1972), for instance, assumes proportional hazards over time; this assumption is not tenable all the time in clinical practice, especially in circumstances with treatment effects that vary with time, and it does not give any direct estimation for the baseline hazard. The Kaplan–Meier estimator Kaplan and Meier (1958) remains the most robust non-parametric estimator of survival functions; however, it is limited to univariate analysis, thereby unable to adjust for any confounding covariate that would have decreased its utility in an inherently heterogeneous population of patients. Parametric survival models require that an investigator assume a specific distribution (for example, exponential, Weibull, or log-normal), and a poor choice of distribution can lead to biased estimates and erroneous inferences.

Data and Evaluation Criteria

Data

It is a dataset of patients of breast cancer which is taken from the SEER program of the National Cancer Institute e.g. November update of the year 2017. The program provides such extensive statistics for cancers based on population data. It denotes the numbers for women diagnosed as having invasive breast cancer in the years 2000 to 2017. The associated records shown are of a patient's age, race, ethnicity, cancer, stage of cancer, size, and treatment grade to them in-a-broad-catchment of statics Teng (2019). for a few seconds. For analysis, we used a dataset available on Kaggle accessed through this link: Breast Cancer ML Notebook. It provides a well-focused picture of class hence imprints patient demographics like age, race, ethnicity, and clinical hematomorphology, such as stage of disease, size, and grade of tumor, along with the treatments given. This is followed by

Evaluation Criteria

Model selection for clinical data generally depends on Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) Bu¨tepage et al. (2022). Developed by Hirotugu Akaike in 1974, the AIC assesses a model's goodness of fit while penalizing complexity by adding a term of 2k, where k is the number of model parameters Akaike (1974). The AIC is given by

$$AIC = -2 \cdot ln(\hat{L}) + 2k, \qquad (8)$$

with Lˆ representing the maximum likelihood of the fitted model. In contrast, the BIC applies a harsher penalty, using k · ln(n) (with n as the sample size) in its formula

$$BIC = -2 \cdot ln(\hat{L}) + k \cdot ln(n). \qquad (9)$$

This further increased penalty made BIC much more conservative than AIC and makes it keep preferring simpler and hence more interpretable models in throw with greater data. Most often AIC is selected for improving predictive accuracy while, for simplicity and clarity, BIC is chosen. Both criteria relate model-fitted complexity, with divergent selection outcomes for models.

Results and Discussion

Kaplan-Meier survival analysis of 4,024 breast cancer patients indicates important survival features figure 1a and 2b. Of these patients, 3,408 (85%) were events during the follow-up study period, which means breast cancer has enormous significance for survival. The most reliable estimation of median survival time was 78 months (95% CI: 77-79 months), which is equally explained in this article prepared accordingly table 1. This analysis is an important benchmark for understanding patient outcomes that could serve comparative studies, treatment assessments, and prognostic models.

Table 1: Kaplan-Meier survival

N	Events	Median	0.95LCL	0.95UCL
4024	3408	78	77	79

The Kaplan-Meier survival analysis of 4,024 breast cancer patients has for long revealed a marked decrease in survival probabilities with time. The survival rate in table 2 shows that survival remains at its best initially, with a survival probability of 99.98% at 1 month and

99.5% at 10 months. Conversely, survival steadily declines with passing time, with 92.3% surviving at 50 months and 77.85% at 60 months. Five years later, the survival rate decreases further to 61.47% at 70 months and 45.84% at 80 months. The median survival time is

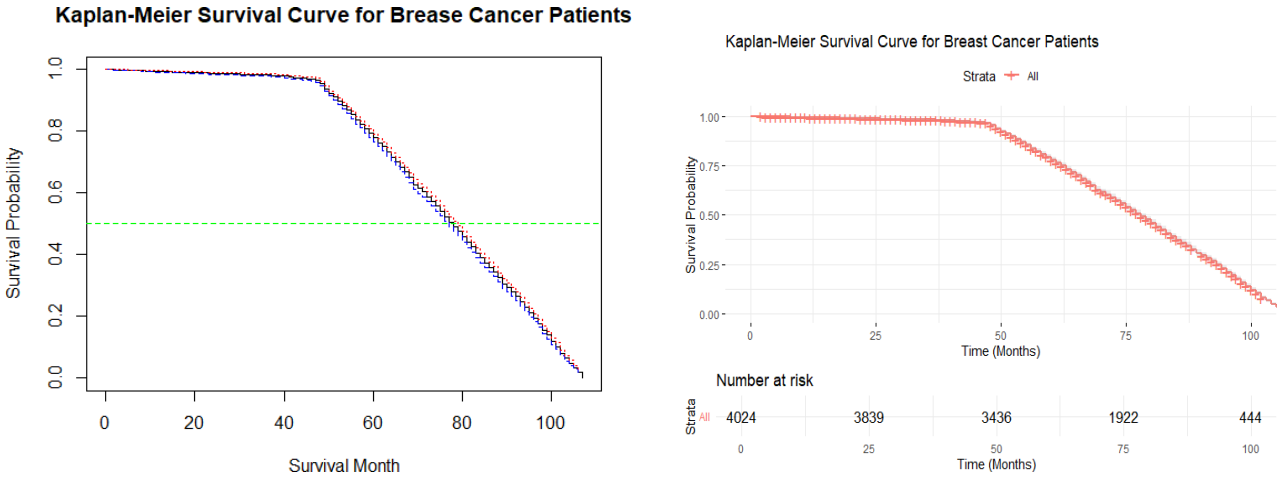

Figure 1: Kaplan-Meier Survival Analysis

approximately 78 months, with a 95 % confidence interval of 77 to 79 months. Subsequently, with 29.46% surviving at 90 months and only 11.97% at 100 months, there was a steeper decline in survival rates. The survival probability rests at zero from 107 months into table 2. This analysis lays bare aspects of extreme significance regarding the long-term prognosis of breast cancer patients and indicates feasibility for putting in place targeted *interventions for improving patient outcome.*

(a) Kaplan-Meier Survival Curve (Overall)
(b) Kaplan-Meier Survival Curve (Heatmap)

The factors influencing survival time in breast cancer patients are examined by the evaluation of several clinical and demographic factors in the univariate Cox proportional hazards analysis and discussed in table 3. Of the factors analyzed, tumor differentiation ($p = 0.0057$), tumor grade ($p = 0.0059$), estrogen receptor status ($p = 0.0039$), and progesterone receptor status ($p = 0.0001$) show statistically significant associations with survival. In poorly differentiated tumor HR = 1.079 connotes a poor prognosis, since aggressive tumors are often associated with rapid disease progression, while a higher tumor grade HR = 0.9269 seems to favor survival when confounding factors may play a role. Whereas in some instances, hormone receptor-positive tumors may be associated with distinct disease progression pattern or resistance to treatment, in this study contrarily estrogen receptor-positive tumors (HR = 1.2632) and progesterone receptor-positive tumors (HR = 1.2155) were assessed to correlate with poorer survival.

Table 2: Kaplan-Meier Survival Estimates for Breast Cancer Patients

Time	n.risk	n.event	Survival	Std. Err	Lower 95% CI	Upper 95% CI
1	4024	1	0.9998	0.000248	0.9993	1
2	4023	1	0.9995	0.000351	0.9988	1
3	4020	2	0.999	0.000497	0.998	1
4	4016	1	0.9988	0.000556	0.9977	0.9998
5	4006	2	0.9983	0.000658	0.997	0.9995
7	3992	1	0.998	0.000703	0.9966	0.9994
8	3987	5	0.9968	0.000898	0.995	0.9985
9	3980	5	0.9955	0.001057	0.9934	0.9976
10	3971	2	0.995	0.001115	0.9928	0.9972
.	.	.	.	.	.	.
.	.	.	.	.	.	.
106	108	47	0.0195	0.002458	0.0152	0.0249
107	61	61	0	NaN	NA	NA

Lymph node involvement or not (N Stage, $p = 0.0869$) showed a near-significant association, in other words, with a possible trend of greater lymph node metastases leading to reduced survival. The number of examined regional lymph nodes and the number of positive regional lymph nodes did not significantly affect the survival rate ($p = 0.05$). These results suggest that tumor differentiation and hormone receptor status are of great importance in predicting breast cancer prognosis, while traditional staging measures may need further evaluation in a multivariate setting to determine their independent effect on survival.

Hazard ratios (HRs) calculated by Cox regression model make a comparison of clinical characteristics with their respective effect by time-to-event outcomes. The dotted line red denotes HR=1 on the horizontal axis indicating no effect. Figure 2a, 2b. Those lying on the right of this line indicate increased risk of the event and those on the left indicate decreased risk. This figure suggests progesterone status and tumor grade were positively associated with hazards indicating increased risk (HR=1) as indicated by blue dots and confidence intervals that do not cross HR=1. Alternatively, estrogen status, differentiation, regional node examination, tumor size, age, regional node positivity, race, T stage, stage 6, N stage, and A stage remained without association to hazard. Their confidence intervals included HR=1, and are represented by black dots. Thus, the only significant predictor for this model about the timing of the event was progesterone status and tumor grade.

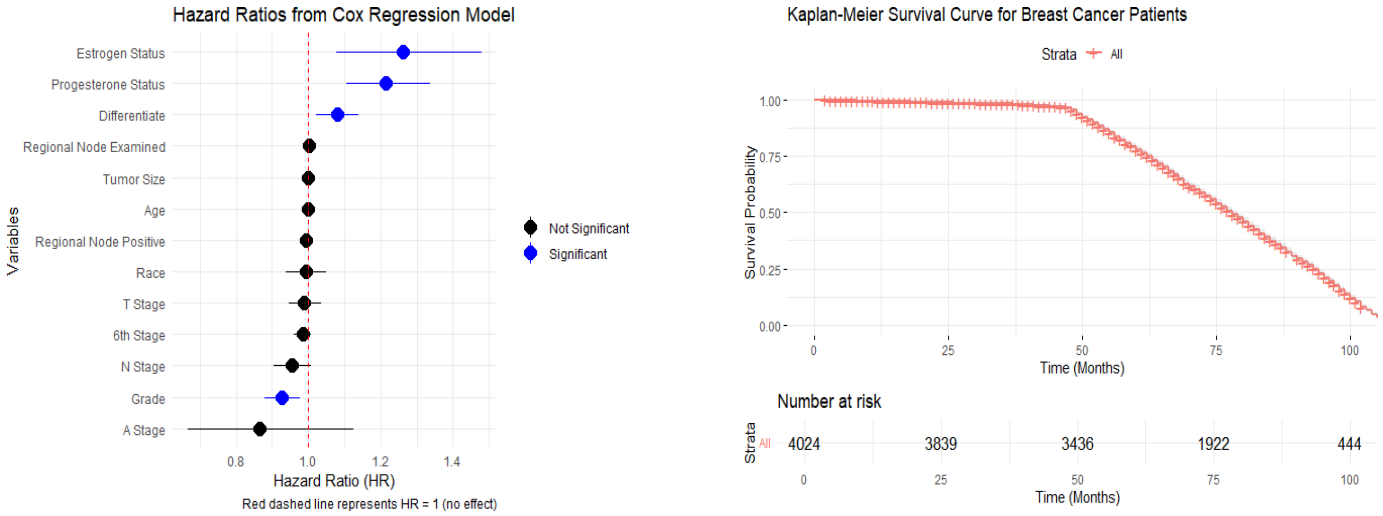

Figure 2: Kaplan-Meier Survival Analysis

(a) Hazard Ratios from univariate Cox Regression Model
(b) Kaplan-Meier Survival Curve (Heatmap)

Overall Cox Proportional Hazards model outputs from Table 4 indicate that the majority of variables, including age, race, tumor stage, and nodal stage, do not appear to contribute significantly toward survival, as evidenced by their p-values remaining high. In contrast, the differentiation status ($p = 0.04563$, HR $= 1.0587$) and progesterone status ($p = 0.00372$, HR $= 1.1716$) emerge as important predictors. This suggests that good differentiation predicts for survival, while increased progesterone possibly elevates risk.

Table 3: Univariate Cox Proportional Hazards Model Results

Variable	Coef	Exp(coef)	Se(coef)	Z	p-value
Age	-0.00012	0.999878	0.00197	-0.062	0.951
Tumor Size	1.48E-05	1.0000	0.000844	0.018	0.986
Race	-0.00697	0.993053	0.028982	-0.241	0.81
T Stage	-0.01045	0.98961	0.02277	-0.459	0.646
N Stage	-0.04614	0.95491	0.02695	-1.712	0.0869
6th Stage	-0.01603	0.9841	0.01205	-1.331	0.183
Differentiate	0.07624	1.07922	0.02756	2.766	0.00568
Grade	-0.07591	0.9269	0.02758	-2.752	0.00592
A Stage	-0.1444	0.8655	0.1337	-1.081	0.28
Estrogen Status	0.23364	1.26318	0.08083	2.891	0.00385
Progesterone Status	0.19518	1.21553	0.04875	4.004	6.23E-05
Regional Node Examined	0.000857	1.000858	0.00215	0.399	0.69
Regional Node Positive	-0.00618	0.993839	0.003932	-1.572	0.116

The likelihood ratio test also gives a p-value of 0.01324 establishing overall significance of the model. But indeed, certain missing values of some variables, e.g. Grade, will muddle analysis.

Table 4: Cox Proportional Hazards Model Results

Variable	Coef	Exp(coef)	Se(coef)	Z	p-value
Age	-0.000	0.9994253	0.0020566	-0.280	0.77983
Race	-0.002	0.9983641	0.0291400	-0.056	0.95519
T Stage	-0.029	0.9710620	0.0480210	-0.612	0.54087
N Stage	-0.028	0.9723141	0.0526061	-0.534	0.59354
6th Stage	0.003	1.0030497	0.0201858	0.151	0.88009
Differentiate	0.057	1.0587138	0.0285438	1.999	0.04563
Grade	NA	NA	0.0000000	NA	NA
A Stage	-0.062	0.9391242	0.1425507	-0.441	0.65950
Tumor Size	0.002	1.0015828	0.0015890	0.995	0.31957
Estrogen Status	0.095	1.0997141	0.0909288	1.045	0.29587
Progesterone Status	0.158	1.1715649	0.0545890	2.901	0.00372
Regional Node Examined	0.002	1.0027231	0.0023235	1.170	0.24184
Regional Node Positive	-0.00	0.9959572	0.0078611	-0.515	0.60633
Marital Status (Married)	0.055	1.0570804	0.0546962	1.015	0.31016
Marital Status (Separated)	0.294	1.3420310	0.1903413	1.546	0.12221
Marital Status (Single)	0.011	1.011	0.0674363	0.173	0.86246
Marital Status (Widowed)	0.033	1.034	0.0902525	0.375	0.70734

The survival data are best fitted with the Weibull distribution according to the model comparisons. The use of AIC and BIC criteria reflects this choice because Weibull has the lowest AIC (30199.01) and BIC (30236.81) (See Table 5) which indicates an appropriate trade-off between goodness-of-fit and complexity in the model. Close in second is the log-logistic model as indicated by its slightly higher AIC and BIC values: 30887.68 and 30925.48, respectively. On the other hand, the exponential model has really poor performance, with retaining the record for highest AIC (37018.88) and BIC (37050.38), thus showing the very poor fit. The degree of importance of the variables is different in the different models.

Table 5: Comparison of Survival Models

Model	Log-Likelihood	AIC	BIC	Significant Variables
Weibull	-15093.5	30199.01	30236.81	Estrogen Status (p = 0.025)
Exponential	-18504.4	37018.88	37050.38	N Stage (p = 0.00076), Estrogen Status (p = 0.029)
Lognormal	-15970.3	31952.67	31990.47	Age (p = 6.3e-5), Estrogen Status (p = 0.012)
Loglogistic	-15437.8	30887.68	30925.48	Estrogen Status (p = 0.011)

The estrogen status is significant on all models and is a strong predictor for survival time. The age is significant in the log-normal model only; possibly its impact on survival time is better expressed under the assumptions of this specific distribution. N Stage is significant for the first

model for exponential death rate only which thus means it has more powerful effect when survival time is modeled using the exponential distribution. However, as the best-fitting model overall for the data was the Weibull model as shown in Figure 3, which has the lowest value on both AIC and BIC, it is therefore the most appropriate modeling for this data sets Survival time. Not all of these variables would be significant under every model, but the Weibull model was flexible enough to fit maximum types of hazard rates so that it would be the best representative of capturing the underlying survival patterns.

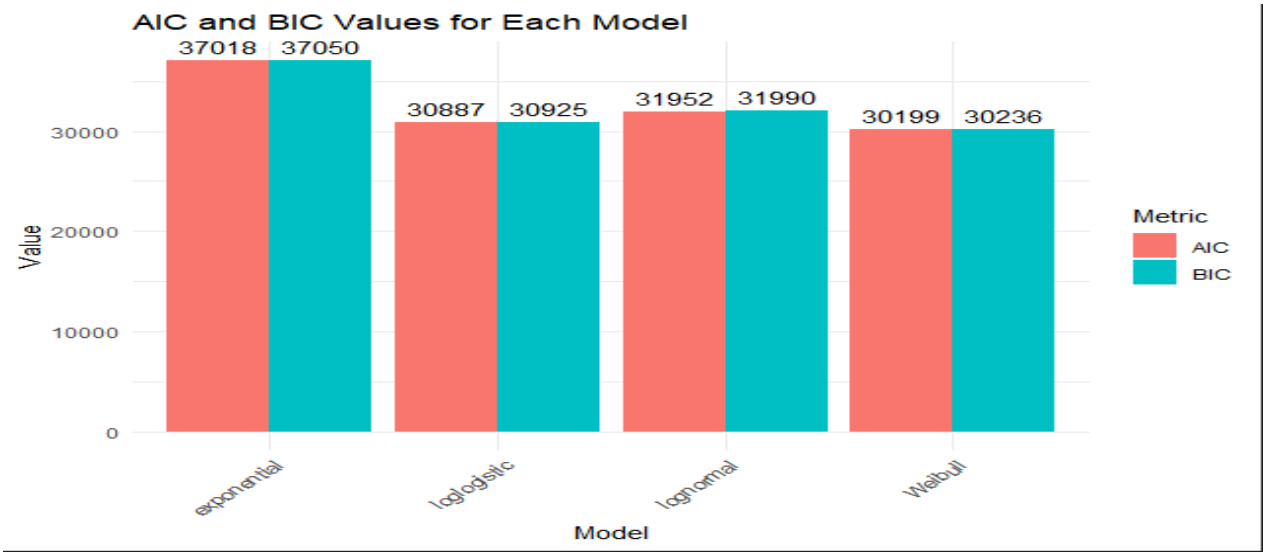

Figure 3: Comparison of AIC and BIC Parametric model

Discussion

The results of this study provide insights into the factors influencing survival outcomes, as determined through model selection using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model with the lowest AIC and BIC values was chosen as the best-fitting model, ensuring a balance between complexity and explanatory power. The Cox proportional hazards model identified key covariates affecting survival, though only a subset of these variables demonstrated statistical significance. This suggests that while multiple factors may contribute to survival, only a few have a pronounced impact when adjusted for other variables. Furthermore, the Kaplan-Meier survival analysis illustrated survival probabilities over time, revealing a marked decline beyond 60 months. This finding underscores the importance of long-term monitoring and intervention strategies to improve survival rates. The consistency between the Kaplan-Meier estimates and the Cox model strengthens the robustness of our findings. However, despite the advantages of the Cox model, potential limitations such as proportional hazards assumption violations and unmeasured confounding factors should be considered. Overall, this study highlights the necessity of using a parsimonious yet informative model that balances interpretability and predictive accuracy. Future research should explore additional variables and alternative modeling technique

Conclusion

This study investigated the relationship between handwriting characteristics and the presence of neurological disorders using a case-control study design. By analyzing various handwriting

features across different tasks, we identified key variables that significantly differentiate individuals with neurological disorders from healthy controls. Our findings suggest that features such as writing pressure, execution time, and jerk movements play a crucial role in distinguishing between the two groups.

The results demonstrate the potential of handwriting analysis as a non-invasive tool for early detection and diagnosis of neurological disorders. The predictive models developed in this study achieved promising classification performance, highlighting the feasibility of using handwriting biometrics for medical screening. However, limitations such as sample size constraints and potential confounding factors should be considered. Future research should explore larger datasets and incorporate machine learning techniques to enhance predictive accuracy.

Overall, this study contributes to the growing field of digital health analytics, emphasizing the importance of handwriting analysis in neurological assessment. The insights gained from this research pave the way for developing automated diagnostic tools that can assist clinicians in early detection and intervention for neurological disorders.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Blanche, P., Dartigues, J. F., & Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine, 32*(30), 5381–5397. https://doi.org/10.1002/sim.5958

Bland, J. M., & Altman, D. G. (1998). Survival probability (the Kaplan-Meier method). *BMJ, 317*(7172), 1572–1580. https://doi.org/10.1136/bmj.317.7172.1572

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Bu¨tepage, G., Vitor, C., & Carlqvist, P. (2022). MSR71 Performance of AIC and BIC for the extrapolation of survival data with different levels of censoring. *Value in Health, 25*(12), S363.

Collett, D. (2015). *Modelling survival data in medical research* (3rd ed.). CRC Press.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187–220. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Ferlay, J., Colombet, M., Soerjomataram, I., et al. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer, 149*(4), 778–789. https://doi.org/10.1002/ijc.33588

Gelman, A., Carlin, J. B., Stern, H., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian survival analysis*. Springer.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457–481. https://doi.org/10.1080/01621459.1958.10501452

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data* (2nd ed.). John Wiley & Sons.

Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data*. Wiley.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., & Mantel, N. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient: II. Analysis and examples. *British Journal of Cancer, 35*(1), 1–39. https://doi.org/10.1038/bjc.1977.1