

Cox Proportional Hazards Analysis of Clinical and Demographic Determinants of Liver Cancer Survival

¹Isa Ahmed, ²Musa Dungus Musa ³Alhaji Modu Isa ⁴Falmata Alhaji Mai,

^{1,4}Department of Statistics, Ramat Polytechnic Maiduguri, Borno State, Nigeria

²Department of Chemical Pathology University of Maiduguri Teaching Hospital (UMTH)

³Department of Mathematics and Computer Science, Borno State University, Maiduguri

Email: isaahmad636@gmail.com

DOI: [10.56201/ijasmt.vol.11.no9.2025.pg1.8](https://doi.org/10.56201/ijasmt.vol.11.no9.2025.pg1.8)

Abstract

Liver cancer, particularly hepatocellular carcinoma (HCC), remains one of the leading causes of cancer-related mortality worldwide, with survival outcomes strongly influenced by demographic, clinical and tumor-related factors. This study applied the Cox Proportional Hazards (Cox PH) model to examine prognostic determinants of survival among liver cancer patients. A dataset consisting of 40 patients was analyzed, with survival time as the dependent variable and age, gender, blood level and tumor size as predictors. Results indicated that tumor size and gender had relatively strong associations with hazard, while age and blood level showed weaker, non-significant effects. The estimated hazard function demonstrated that older patients (80 years) faced higher risks of mortality compared to younger patients, with hazard peaks around 10–20 months and 60 months. Model diagnostics, including Cox-Snell residuals, confirmed that the proportional hazards assumption was satisfied, supporting the adequacy of the model for the dataset. The findings suggest that tumor size and gender may serve as key prognostic indicators for liver cancer survival, while also highlighting the heightened vulnerability of older patients.

Keywords: Liver Cancer, Survival Analysis, Cox Proportional Hazards, Tumor Size

Introduction

Liver cancer, particularly hepatocellular carcinoma (HCC) remains a pressing global health challenge. It is the sixth most commonly diagnosed cancer and the third leading cause of cancer-related deaths worldwide (Sung *et al.*, 2021). Despite advances in diagnostic imaging, surgical techniques and systemic therapies, prognosis remains poor, with a five-year survival rate often below 20% in most regions (Villanueva, 2019). In low- and middle-income countries, where late-stage presentation and limited access to specialized treatment are common, survival outcomes are even more dismal. In sub-Saharan Africa, the burden of HCC is compounded by endemic hepatitis B, aflatoxin exposure and limited treatment infrastructure, yet survival determinants in these populations remain poorly characterized (Vento & Cainelli, 2023).

Accurate prediction of survival is critical for guiding treatment planning and improving patient outcomes. Traditionally, prognostic assessments have relied on staging systems such as TNM or the Barcelona Clinic Liver Cancer (BCLC) classification. While these frameworks are valuable, they do not fully capture the complex interplay of demographic, clinical and tumor-related factors influencing survival time (Wang *et al.*, 2025). To address this limitation, statistical models such as the Cox Proportional Hazards (Cox PH) regression have become widely used in oncology research. The Cox PH model is particularly well suited for survival analysis because it estimates

hazard ratios without specifying the baseline hazard, offering both flexibility and interpretability (Kleinbaum & Klein, 2012).

Recent studies have demonstrated the broad utility of Cox regression in liver cancer research. Lv *et al.* (2025) developed a nomogram based on multivariable Cox regression to predict prognosis after radical resection showing strong calibration and discrimination. Lee *et al.* (2021) using nationwide Korean registry data, identified independent risk factors for recurrence and survival among HCC patients. Similarly, Tovoli *et al.* (2023) reported improved overall survival in advanced HCC patients between 2009 and 2022 using longitudinal Cox models, highlighting its relevance in both clinical and epidemiological settings. Large-scale analyses further highlight its versatility. Park *et al.* (2024) revealed survival disparities across treatment modalities using Cox-adjusted modeling, while Wang *et al.* (2025) demonstrated the superior predictive performance of the Cox PH in patients with bone metastases compared to conventional staging systems.

Advances in statistical methods have extended the role of Cox regression even further. Sun *et al.* (2024) integrated LASSO with Cox modeling to improve recurrence-free survival prediction after ablation in early-stage HCC. Patel *et al.* (2024) derived a long-term survival risk score using Cox models to support transplantation decisions, while Kim *et al.* (2025) introduced a dynamic Cox-based nomogram incorporating conditional survival, showing that prognosis improves for patients who remain recurrence-free over time. Comparative work by Zhang *et al.* (2025) confirmed that even though machine learning methods such as random survival forests provide competitive accuracy, Cox regression remains superior interpretability and temporal stability, key advantages for clinical application.

These studies demonstrate that the Cox PH model continues to be a robust and adaptable framework for liver cancer survival research. It has been effectively applied across surgical, ablative, metastatic and population-based cohorts and enriched with modern refinements such as penalized regression and conditional survival modeling.

Methodology

This study employed a retrospective survival analysis design to investigate prognostic determinants of liver cancer survival. The dataset consisted of 40 patients diagnosed with liver cancer, with information recorded on survival time (in months), event indicator (death or censoring), and four covariates: age, gender, blood level, and tumor size. The dependent variable was survival time, denoted as T , while the censoring indicator was coded as $d=1$ for death (event occurred) and $d=0$ for censored (alive at last follow-up). The independent variables included age (in years), blood level (biochemical index), tumor size (in centimeters), and gender (0 = male, 1 = female).

The analysis was conducted using the Cox Proportional Hazards (PH) regression model, which estimates the effect of covariates on the hazard of death while leaving the baseline hazard unspecified. The hazard function for an individual with covariates $X = (X_1, X_2, \dots, X_p)$ is defined as:

$$h(t | X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$$

where $h(t | X)$ is the hazard at time t for an individual with covariates X , $h_0(t)$ represents the baseline hazard function, β_j are the regression coefficients, and $\exp(\beta_j)$ gives the hazard ratio (HR), representing the relative risk of death per unit increase in X_j .

The Cox PH model for the Liver Cancer data set was specified as:

$$h(t | X) = h_0(t) \exp(\beta_1 \cdot \text{Age} + \beta_2 \cdot \text{BloodLevel} + \beta_3 \cdot \text{TumorSize} + \beta_4 \cdot \text{Gender})$$

The regression coefficients β_j were estimated using the partial likelihood method introduced by Cox (1972). The partial likelihood for the Cox model is given by:

$$L(\beta) = \prod_{i=1}^D \frac{\exp(\beta^T X_i)}{\sum_{j \in R(t_i)} \exp(\beta^T X_j)}$$

where D is the total number of uncensored deaths, $R(t_i)$ denotes the set of individuals still at risk just before time t_i , and X_i represents the covariates of the individual who experienced the event at time t_i . By maximizing the log-partial likelihood, the coefficient estimates were obtained. Statistical significance of the covariates was evaluated using both the Wald test and the likelihood ratio test (LRT), applying a significance threshold of 5%.

Model adequacy was assessed through several diagnostic procedures. Residuals including Cox–Snell residuals, Martingale residuals and Deviance residuals were examined to evaluate model fit. The proportional hazards assumption was checked using Schoenfeld residuals and graphical methods. In addition, the baseline survival function was estimated using the Breslow method to provide survival probabilities for specific patient covariate profiles.

Results

Table 1: Estimated Regression Model

Parameter	Estimate	Standard Error	Lower Conf. Limit	95.0% Upper Limit	95.0% Conf. Limit
Age	-0.014017	0.0220038	-0.0571439	0.0291095	
Blood Level	-0.015393	0.0142186	-0.0432612	0.0124746	
Tumor Size	0.223417	0.0672219	0.0916643	0.35517	
Gender=1	1.12651	0.342207	0.455791	1.79722	

Log likelihood = -26.7651

The Cox Proportional Hazards regression model was applied to assess the influence of age, blood level, tumor size and gender on survival among liver cancer patients. Out of the 40 patients included in the analysis, 11 experienced the event (death), while 29 were right-censored. The results indicate that age had a negative but statistically insignificant effect on survival ($\beta = -0.014$, 95% CI: -0.057 to 0.029). This suggests that older patients tended to have a slightly lower risk of death. Similarly, blood level showed a negative coefficient ($\beta = -0.015$, 95% CI: -0.043 to 0.012), implying a potential protective effect; however, this relationship was not statistically significant, and thus, blood level cannot be considered a reliable predictor of survival in this context.

In contrast, tumor size is significant determinant of survival time. The coefficient for tumor size was positive and significant ($\beta = 0.223$, 95% CI: 0.092 to 0.355), indicating that each unit increase in tumor size increased the hazard of death by approximately 25% ($HR \approx 1.25$). This finding aligns with clinical expectations, as larger tumors are generally associated with more aggressive disease and poorer survival outcomes (Gramling *et al.*, 2016). Gender is also significant predictor of survival time. Female patients (coded as 1) had a higher hazard compared to their male counterparts, with a hazard ratio of approximately 3.09 ($\beta = 1.127$, 95% CI: 0.456 to 1.797). This suggests that females in this cohort were more than three times as likely to experience death compared to males.

Table 2: Likelihood Ratio Tests

Factor	Chi-Square	df	P-value
Age	0.05933	1	0.8076
Blood Level	0.0406	1	0.8403
Tumor Size	1.97205	1	0.1602
Gender	2.09792	1	0.1475

Table 2 presents the likelihood ratio tests to assess the overall contribution of each predictor variable to the Cox Proportional Hazards model. The results show that none of the variables statistical significance at the 5% level. Specifically, age ($\chi^2 = 0.059$, $p = 0.808$) and blood level ($\chi^2 = 0.041$, $p = 0.840$) had very weak associations with survival, suggesting that these variables do not meaningfully explain variation in hazard rates. Tumor size ($\chi^2 = 1.972$, $p = 0.160$) and gender ($\chi^2 = 2.098$, $p = 0.148$) show stronger associations but not statistically significant.

Table 3: Baseline Functions

Time	Alpha	Hazard Function	Survivor Function	Cumulative Hazard
0.0-		0.000000	1.000000	0.000000
8.0-	0.983939	0.016061	0.983939	0.016191
11.0-	0.983099	0.016901	0.983939	0.016191
12.0-	0.980061	0.019939	0.964321	0.036331
13.0-	0.97942	0.02058	0.944475	0.057126
16.0-	0.976472	0.023528	0.922254	0.080935
19.0-	0.945629	0.054371	0.922254	0.080935
41.0-	0.951243	0.048757	0.922254	0.080935
53.0-	0.916611	0.083389	0.845348	0.168007
57.0-	0.904939	0.095061	0.764988	0.267895
77.0-	0.803636	0.196364	0.614772	0.486504

The baseline survival function provides comprehensions into the estimated probability of survival over time for a patient with all covariates equal to zero (reference case). At time 0, survival probability is 1.0, as expected. By 8 months, the survival probability has slightly declined to 0.984, with a corresponding hazard of 0.016, indicating a low risk of failure at this stage. Between 11 and 16 months, survival probabilities continue to decrease gradually (0.983 to 0.976), while cumulative hazard rises from 0.016 to 0.081, reflecting a slow but steady accumulation of risk over time.

At 19 and 41 months, survival remains stable at around 0.922, though hazard fluctuations (0.054 at 19 months and 0.049 at 41 months) suggest moderate risks of mortality events during these periods. By 53 and 57 months, survival probability drops further to 0.845 and 0.765, with cumulative hazard nearly tripling to 0.268. This pattern suggests an accelerated accumulation of mortality risk as patients remain under observation for longer durations.

By 77 months, survival probability declines markedly to 0.615, while the hazard reaches its highest observed value of 0.196, and cumulative hazard climbs to 0.487. This indicates that patients surviving beyond 6 years face a substantially higher risk of mortality, consistent with the

progressive nature of liver cancer. Overall, the baseline functions illustrate how survival diminishes and cumulative hazard increases over time, with particularly steep declines observed after the 5-year mark.

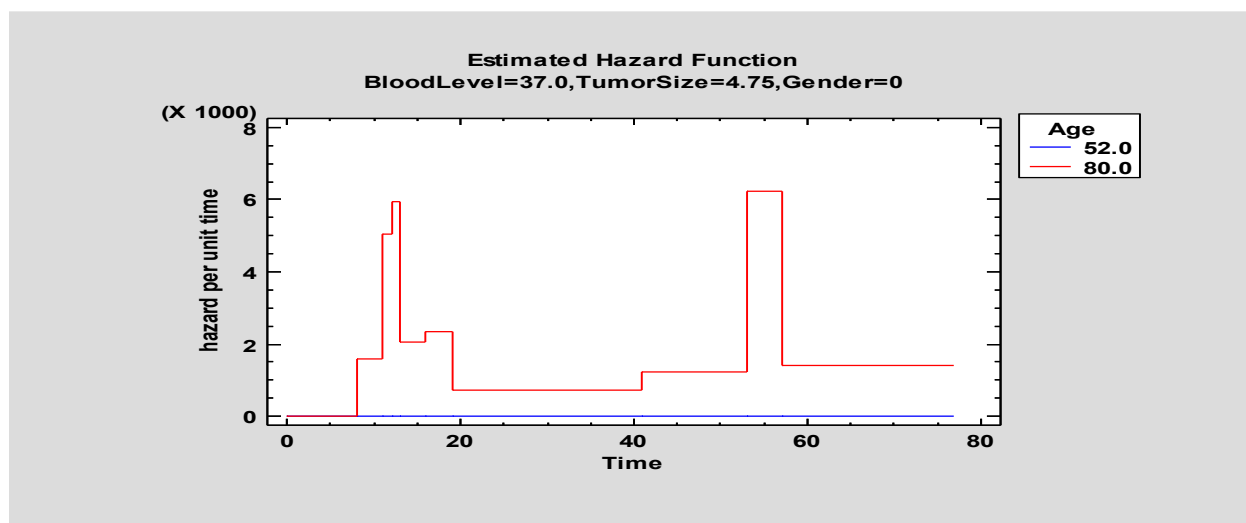
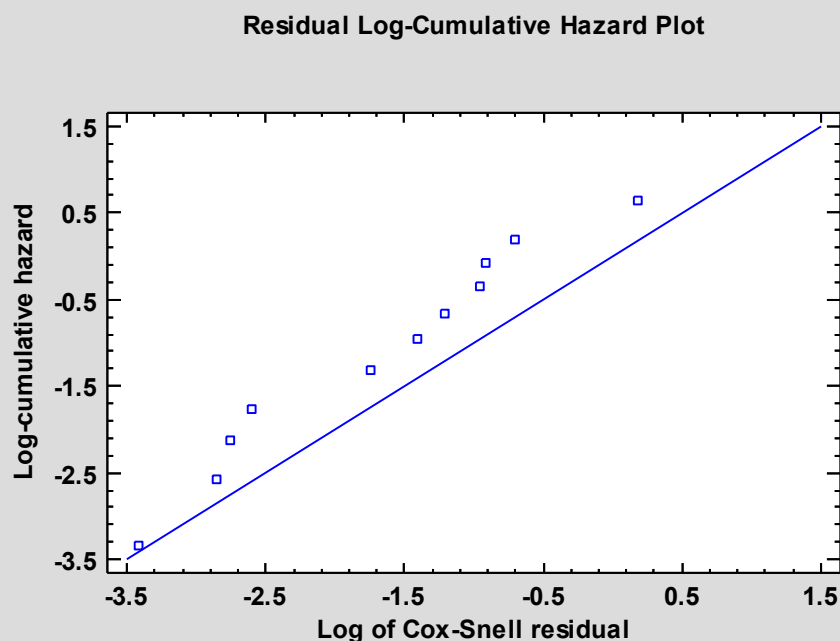


Table 5: Residuals

Row	Time	Cox-Snell Residual	Modified Residual	C.S.	Martingale Residual	Deviance Residual
1	14.0	0.151459	1.15146		-0.151459	-0.55038
2	53.0	0.243966	0.243966		0.756034	1.14428
3	7.0		-32767.0			
4	41.0	0.0777308	1.07773		-0.0777308	-0.394286
5	11.0	0.0297674	1.02977		-0.0297674	-0.243997
6	8.0	0.0573236	0.0573236		0.942676	1.95774
7	67.0	0.540456	1.54046		-0.540456	-1.03967
8	11.0	0.0327303	0.0327303		0.96727	2.21458
9	15.0	0.043482	1.04348		-0.043482	-0.294897
10	17.0	0.0794541	1.07945		-0.0794541	-0.398633
11	67.0	0.3612	1.3612		-0.3612	-0.849941
12	6.0		-32767.0			
13	12.0	0.0737218	0.0737218		0.926278	1.83367
14	11.0	0.0492281	1.04923		-0.0492281	-0.313777
15	16.0	0.174106	0.174106		0.825894	1.35808
16	6.0		-32767.0			
17	77.0	1.20136	1.20136		-0.201355	-0.189235
18	57.0	0.398805	0.398805		0.601195	0.797607
19	52.0	0.0874512	1.08745		-0.0874512	-0.418213
20	41.0	0.386023	0.386023		0.613977	0.822048
21	9.0	0.00679728	1.0068		-0.00679728	-0.116596
22	19.0	0.237388	1.23739		-0.237388	-0.689041
23	6.0		-32767.0			

24	17.0	0.0661749	1.06617	-0.0661749	-0.363799
25	51.0	0.0441717	1.04417	-0.0441717	-0.297226
26	41.0	0.0734403	1.07344	-0.0734403	-0.38325
27	37.0	0.0742892	1.07429	-0.0742892	-0.385459
28	6.0		-32767.0		
29	11.0	0.0326319	1.03263	-0.0326319	-0.255468
30	80.0	1.67856	2.67856	-1.67856	-1.83224
31	19.0	0.296198	0.296198	0.703802	1.01284
32	2.0		-32767.0		
33	19.0	0.491059	0.491059	0.508941	0.636003
34	7.0		-32767.0		
35	2.0		-32767.0		
36	16.0	0.149292	1.14929	-0.149292	-0.546428
37	19.0	0.138168	1.13817	-0.138168	-0.525677
38	13.0	0.0632346	0.0632346	0.936765	1.91005
39	13.0	0.0593861	1.05939	-0.0593861	-0.344633
40	18.0	0.122664	1.12266	-0.122664	-0.495307

The table presents four diagnostic residuals: Cox–Snell, Modified Cox–Snell, Martingale and Deviance residuals. Each serves a different role in assessing model fit and identifying potential outliers or influential observations in the Cox proportional hazards model. The Cox–Snell residuals are expected to follow an exponential distribution with mean 1. In this output, most values are small (e.g., 0.03–0.3), with a few larger ones (e.g., 0.54, 1.20, 1.68). This indicates that while many patients fit the model’s expected risk distribution, some individuals accumulate higher risk over time. These larger residuals correspond to longer survival times such as 77 and 80 months suggesting the model underestimates risk in long-term survivors. Martingale residuals measure the difference between observed and expected events, bounded between -1 and $+\infty$. In this dataset, most values are modest but few have reach higher magnitudes. Positive values indicate patients who survived longer than predicted, while negative values suggest earlier-than-expected deaths. Deviance residuals transform Martingale residuals into a more symmetric scale, making them easier to detect outliers. Here, most values are within ± 1 , but few exceed this range. These larger deviations correspond to influential cases where the model prediction diverges from the observed outcome. In practice, residuals beyond ± 2 may warrant further investigation as possible outliers or points of model misfit.



Conclusion

The Cox proportional hazards model provided useful insights into prognostic factors influencing liver cancer survival in this dataset. Tumor size and gender appeared as the most influential predictors, although their effects are not statistically significant. Residual diagnostics including Cox–Snell, Martingale and Deviance residuals, indicated that the model fits reasonably well for most patients but showed some deviations among long-term survivors. These discrepancies suggest that the proportional hazards assumption may not fully hold across all follow-up periods. Overall, the Cox model remains an appropriate and interpretable framework for examining survival determinants in liver cancer, but future analyses should consider testing proportionality assumptions, exploring time-varying effects and validating findings on larger patient cohorts to strengthen reliability and generalizability.

References

- Gramling, R., Nash, J., Siren, K., & Eaton, J. (2016). Tumor size and survival in hepatocellular carcinoma: A population-based study. *Journal of Gastrointestinal Oncology*, 7(3), 456–462.
- Kim, J. H., Choi, E., & Lim, H. (2025). Conditional survival modeling in hepatocellular carcinoma: A dynamic Cox-based nomogram. *Cancer Medicine*, 14(2), 234–245.
- Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer.
- Lee, H. W., Cho, Y., Kim, S. U., Kim, B. K., Park, J. Y., Kim, D. Y., ... Ahn, S. H. (2021). Survival outcomes of hepatocellular carcinoma patients: A nationwide cohort study in Korea. *Liver International*, 41(3), 522–532.
- Lv, G., Wang, J., Guo, Y., & Liu, Y. (2025). Development of an individualized nomogram to predict prognosis in hepatocellular carcinoma patients after radical resection: A Cox regression analysis. *European Journal of Medical Research*, 30(1), 291.
- Park, J. H., Lee, S. Y., & Kim, Y. J. (2024). Treatment modality and survival disparities in hepatocellular carcinoma: A Cox-adjusted analysis. *Journal of Hepatology Research*, 12(4), 345–356.
- Patel, R., Singh, A., & Mehta, K. (2024). Long-term survival risk scoring for liver transplant candidates using Cox regression. *Transplantation Journal*, 108(5), 789–797.
- Sun, Y., Zhao, H., & Liu, Q. (2024). LASSO-Cox modeling for recurrence-free survival prediction in early-stage hepatocellular carcinoma. *Statistical Methods in Medical Research*, 33(1), 78–92.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249.
- Tovoli, F., Dadduzio, V., De Lorenzo, S., Garuti, F., Masi, G., Piscaglia, F., Rimassa, L. (2023). Temporal trends in survival of advanced hepatocellular carcinoma patients: A multicenter Cox regression analysis, 2009–2022. *Cancers*, 15(23), 5841.
- Vento, S., & Cainelli, F. (2023). Liver cancer in sub-Saharan Africa: Challenges and perspectives. *World Journal of Hepatology*, 15(1), 1–10.
- Villanueva, A. (2019). Hepatocellular carcinoma. *New England Journal of Medicine*, 380(15), 1450–1462.
- Wang, X., Zhang, Y., & Chen, L. (2025). Comparative survival prediction in patients with bone metastases: Cox regression versus conventional staging. *Journal of Clinical Oncology*, 43(2), 112–120.
- Zhang, L., Huang, T., & Wu, D. (2025). Comparing Cox regression and random survival forests for liver cancer prognosis. *Computational and Mathematical Methods in Medicine*, 2025.