Journal of Quality Measurement and Analysis *JQMA* 22(SI) 2026, 129-144
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.22si.2026.08

COMPARING COX PROPORTIONAL HAZARDS (COX PH) AND LOG-LOGISTIC SURVIVAL REGRESSION (LLSR) MODELS FOR PREDICTING SURVIVAL IN CANCER-ASSOCIATED THROMBOSIS (CAT)

(*Perbandingan antara Model Cox Proportional Hazards (Cox PH) dan Regresi Ketahanan Log-Logistik (LLSR) dalam Meramalkan Kelangsungan Hidup Pesakit Trombosis Berkaitan Kanser (CAT)*)

NUR ADLINA NISA HAJAZI, SITI AFIQAH MUHAMAD JAMIL*,
NORSHAHIDA SHAADAN, LAILATULEMA ABBAS & NOR RAFEAH TUMIAN

ABSTRACT

This study addresses the heightened risk of Cancer-associated Thrombosis (CAT), a significant cause of morbidity and mortality among cancer patients, particularly focusing on venous thromboembolism (VTE). Although VTE risk is perceived as lower in Asian populations, evidence suggests that factors like existence of chemotherapy treatment, advanced cancer staging, and immobility can negate this assumption. This study compares two survival analysis models: Cox Proportional Hazards (Cox PH) and Log-Logistic Survival Regression (LLSR) to evaluate their predictive capabilities in modelling CAT survival. A prospective cohort of 249 adult cancer patients from Hospital Canselor Tuanku Muhriz (HCTM) and Hospital Kuala Lumpur (HKL) was analysed. Survival time was defined as the duration between cancer diagnosis to death or censoring. Predictors included demographic and clinical variables. RStudio was used to fit the Cox PH and LLSR models. Evaluation metrics, including classification accuracy, sensitivity, specificity, and AUROC, are applied to real CAT datasets. Both models were successful in identifying significant survival predictors. The findings underscore the model's effectiveness in predicting CAT risk, thereby advancing the precision of early detection and personalized treatment strategies for cancer patients. The model evaluation showed that Coc PH model performs slightly better compared to LLSR model which may be due to its flexibility to model hazard function without assuming specific distribution. This study's findings are poised to advance public health by optimizing CAT management, leading to improved patient care and clinical decision-making.

Keywords: cancer-associated thrombosis; survival analysis; Cox proportional hazard; log-logistic survival regression

ABSTRAK

Kajian ini menumpukan kepada risiko tinggi Trombosis Berkaitan Kanser (Cancer-associated Thrombosis, CAT), yang merupakan penyumbang utama kepada morbiditi dan mortaliti dalam kalangan pesakit kanser, khususnya melibatkan venous thromboembolism (VTE). Walaupun risiko VTE sering dianggap lebih rendah dalam populasi Asia, bukti menunjukkan bahawa faktor seperti kewujudan rawatan kemoterapi, tahap kanser yang lanjut, dan ketidakmobilan boleh menafikan andaian tersebut. Kajian ini membandingkan dua model analisis ketahanan iaitu Cox Proportional Hazards (Cox PH) dan Regresi Ketahanan Log-Logistik (LLSR) bagi menilai keupayaan ramalan dalam memodelkan ketahanan CAT. Satu kohort prospektif yang melibatkan 249 pesakit kanser dewasa dari Hospital Canselor Tuanku Muhriz (HCTM) dan Hospital Kuala Lumpur (HKL) telah dianalisis. Masa ketahanan ditakrifkan sebagai tempoh antara diagnosis kanser hingga kematian atau penapisan (censoring). Pemboleh ubah peramal merangkumi faktor demografi dan klinikal. Perisian RStudio telah digunakan untuk membina

model Cox PH dan LLSR. Metrik penilaian termasuk ketepatan klasifikasi, sensitiviti, spesifisiti, dan AUROC telah diaplikasikan ke atas data sebenar CAT. Kedua-dua model berjaya mengenal pasti peramal ketahanan yang signifikan. Dapatan kajian ini menekankan keberkesanan model dalam meramalkan risiko CAT, sekali gus meningkatkan ketepatan pengesanan awal dan strategi rawatan yang diperibadikan bagi pesakit kanser. Penilaian model menunjukkan bahawa model Cox PH berprestasi sedikit lebih baik berbanding model LLSR, yang mungkin disebabkan oleh fleksibilitinya dalam memodelkan fungsi hazard tanpa andaian taburan tertentu. Dapatan kajian ini berpotensi untuk memperkukuh kesihatan awam melalui pengoptimuman pengurusan CAT, seterusnya meningkatkan kualiti penjagaan pesakit dan sokongan dalam membuat keputusan klinikal.

Kata kunci: trombosis berkaitan kanser; analisis ketahanan; Cox proportional hazards; regresi ketahanan log-logistik

1. Introduction

Cancer-associated thrombosis (CAT) refers to a formation of abnormal blood clots in cancer patient, a subset of venous thromboembolism (VTE), where the blood clots forms in the venous system (Peng *et al*. 2024). It is a serious concern due to its elevated risk in cancer patients, leading to increase in mortality and morbidity. It is the second leading cause of death among cancer patients, after the cancer progression itself (Donnellan & Khorana 2017). Studies have shown that cancer patients are at higher risk of VTE compared to the general population (Mulder *et al*. 2021; Lee *et al*. 2021).

In the Asian context, there is a growing awareness of the impact of CAT. According to a recent systematic study by Lee *et al*. (2021), the rate of incidence of CAT ranges between 1.85 and 9.88 per 1000 individuals among the Asian population. Specifically, a study conducted in Kuala Lumpur, Malaysia, found that the incidence of CAT in patients with solid organ cancer was 4.8%, particularly those with poor performance status and higher body mass index (Abbas *et al*. 2025). This increased risk of VTE in cancer patients can worsen prognosis and complicates cancer management as there is a greater risk of recurring thrombosis and bleeding complications (Donnellan & Khorana, 2017). CAT does not only impact on the patient survival but also results to significant healthcare costs, as hospitalizations for VTE are significantly more expensive for cancer patients with VTE (Ay *et al*. 2017).

Considering its significant influence on survival and healthcare resources, accurate survival prediction in patients with CAT is therefore crucial for guiding clinical decision-making, resource allocation, and optimizing patient outcomes. Survival analysis aids clinicians to predict the probability of survival and make informed treatment decisions by providing a statistical framework for estimating time-to-event outcomes (Wang *et al*. 2024). The Cox PH model is one of the most used models, and it follows the proportional hazards assumption where the risk of an event remains constant over time (McLernon *et al*. 2023). Despite being commonly utilised, this assumption may not always align with the complex nature of disease progression in CAT patients since the risk of thrombosis and mortality can vary significantly with disease progression and patients undergo different treatments (Zhang *et al*. 2018)

An alternative is the LLSR model, which does not rely on the proportional hazards assumption and allows for non-monotonic hazard functions (Kleinbaum & Klein 1996). This model is particularly helpful in situations when the hazard function is non-monotonic, that is, rising and then falling over time as is often the case with cancer outcomes (Borges 2021). While both models have been widely used and compared in non-CAT cancer contexts (e.g.,

Hashemian *et al*. 2017 in colorectal cancer), there is limited, if any, literature comparing their performance in predicting outcomes among CAT patients.

The aim of this study is to evaluate and compare the Cox PH and LLSR models for predicting survival among CAT patients. This study aims to determine a suitable modelling approach that can assist better prognostic assessment and enhance the management of CAT in real-world settings by comparing and assessing their prediction performance using clinical data from hospitals in Malaysia.

2. Literature Review

2.1. *Introduction*

Accurate survival modelling is essential in medical research, especially for conditions like CAT, where mortality risk is high and varies across individuals. The time until an event of interest, usually death, can be estimated using survival analysis, which also helps determine the factors that affect this outcome. Two widely used survival modelling approaches are the Cox PH model and the LLSR model, each with its own advantages and assumptions.

2.2. *Cancer-associated thrombosis*

Cancer patients face a spectrum of risks that contribute to the development of cancer-associated thrombosis (CAT). For instance, circulatory statis, notably prevalent in the presence of bulky tumors, can lead to local vascular compression and diminished blood flow (Fernandes *et al*. 2019). Surgical interventions, including tumor debulking and excision, further heighten the risk of venous statis due to extended hospitalization. The prevalence of oncologic pain necessitates opioid therapy in 30-50% of patients during active treatment, with this figure surpassing 70% in advanced stages of the disease (Fallon *et al*. 2018). Such pain management can lead to prolonged periods of immobility, compounding the risk factors associated with CAT in the context of cancer patients. Besides, in Malaysia, a cross-sectional study on CAT among solid tumor patients from 2015 to 2019 at National Cancer Institute showed that 5.7% (n=688) patients developed CAT from the 12,093 patients recruited. 48.1% were treated with fondaparinux, 45.1% with enoxaparin, and 6.8% with rivaroxaban. The main finding in this study was that the percentage of health utilization per budget allocation was reported to be 2.1-5.2% (mean 3.3%) of total institution drugs budget and Rivaroxaban was the cheapest cost per patient per year (Carrier *et al*. 2019). In Malaysia, Abbas *et al*. (2025) reported that the incidence of CAT among cancer patients in Kuala Lumpur was 4.8%, with higher risk observed in patients with poor performance status and higher BMI at baseline. This emphasizes the clinical significance of CAT in the Malaysian population and underscores the need for local survival modelling studies, as most existing study comes from other populations.

2.3. *Cox proportional hazards model*

Cox (1972) developed the semi-parametric Cox PH model, which is now a fundamental component of survival analysis in clinical research. It is one of the most used multivariate statistical models in survival analysis, valued for its semi-parametric structure and ability to incorporate multiple covariates without requiring specification of the baseline hazard function (Deo *et al*. 2021).

According to the proportional hazard assumption, the hazard ratios between individuals remain constant over time (McLernon *et al*. 2023). Although this assumption makes interpretation easier, it may not always be true in actual clinical settings. Violations of the

proportional hazards assumption can lead to biased and inaccurate results, particularly in clinical contexts where non-proportional hazards are often observed, such as oncology trials (Bardo *et al*. 2024).

Despite its limitations, the Cox model remains widely used in oncology research due to its interpretability and ease of use. It has been used effectively to identify VTE as a significant prognostic factor influencing mortality in cancer patients (Crobach *et al*. 2023). However, different modelling approaches can be more suitable when the proportional hazards assumption does not hold.

2.4. *Log-logistic survival regression model*

Log-logistic survival regression is a fully parametric model that assumes the logarithm of survival times follows a logistic distribution (Kleinbaum & Klein 1996). This model is particularly helpful in situations when the hazard function is non-monotonic, meaning it rises and then falls over time, as is often the case with cancer outcomes (Borges 2021). In contrast to the Cox model, the log-logistic model provides explicit expressions for both the survival and hazard functions, enabling for extrapolation beyond the observed follow-up period and direct estimation of survival probabilities (Kleinbaum & Klein 1996).

Several studies have highlighted the effectiveness of log-logistic models in oncology. Borges (2021) applied the model to cervical cancer data and explored its ability to model non-monotonic hazard functions and long-term survival. Similarly, Kumar *et al*. (2020) applied a log-logistic model on patients with lung cancer and demonstrated its superior fit for analysing survival data. These applications underscore the advantages of log-logistic models when modelling complex hazard shapes and providing clinically significant interpretation.

However, the traditional log-logistic model assumes that covariate effects on survival are constant over time on the scale of odds of survival (Proportional Odds) or the timing of the event (Accelerated Failure Time), rather than on the hazard rate (Kleinbaum & Klein 1996). This assumption often fails in real-world medical data, particularly in CAT, as variables like VTE can have varying effects on survival over time (Khan *et al*. 2017). Nonetheless, LLSR can be serve as a powerful and interpretable alternative for Cox PH when the hazard structure is non-proportional, but the covariate effects remain constant.

2.5. *Conclusion*

Cancer-associated thrombosis remains a significant contributor to increased morbidity and mortality among cancer patients, yet survival prediction in this context is often constrained by methodological limitations. The commonly used Cox Proportional Hazards (Cox PH) model, while popular for its ease of interpretation, relies on the proportional hazards assumption which may not hold true in dynamic clinical scenarios such as CAT, where risks can vary significantly over time due to disease progression and evolving treatment plans. Although alternative models like Log-Logistic Survival Regression (LLSR) offer greater flexibility for non-monotonic hazard functions, comparative evidence on their predictive performance in CAT is scarce, particularly within Southeast Asian populations.

Given this gap, the present study is justified in its aim to evaluate and compare Cox PH and LLSR models using real-world clinical data from Malaysian hospitals. By doing so, this research not only addresses the methodological gap in survival modelling for CAT but also provides region-specific insights that are crucial for improving prognostic accuracy, guiding treatment strategies, and optimizing resource allocation in oncology care. The findings are expected to contribute directly to evidence-based decision-making and enhance the precision of CAT management, ultimately supporting better patient outcomes in local clinical practice.

3. Methodology

3.1. *Study design and participants*

The study design used is prospective study which involved data from 249 adult patients diagnosed with solid organ cancer who were treated in Hospital Canselor Tuanku Muhriz (HCTM) and Hospital Kuala Lumpur (HKL) between May 2022 until August 2023. Any patients with newly diagnosed or recurrent solid cancer aged 18 years and older who had confirmed histopathological findings of cancer at the time of diagnosis were included and were followed up in the clinic or during hospital admission. Patients diagnosed with a thrombotic event, whether symptomatic or asymptomatic, received standard treatment according to facility guidelines without intervention based on the standard protocol. Patients with haematological malignancies, pregnant patients or those who were diagnosed with a thrombotic event prior to a solid cancer diagnosis were excluded from the studies. This study has been approved by the Medical Research Ethics Committee of Universiti Kebangsaan Malaysia (FF-2022-099) and Medical Research and Ethics Committee (MREC) of the Malaysian Ministry of Health (MOH) (NMRR ID -22-01438-MW4 (IIR)).

3.2. *Outcome and variable definitions*

Survival time was measured from the date of cancer diagnosis until the date of death or last follow-up. Patients who survived at the end of the follow-up period were censored. For the outcome variable, death was coded as 1 and censored cases were coded as 0. Predictor variables included sociodemographic data (age group, gender, smoking status), clinical data (BMI category, ECOG performance status), data on cancer (cancer type, cancer stage, disease status), treatment modalities (surgery, chemotherapy, targeted therapy, hormonal therapy, radiation, or combination), and the presence or absence of cancer-associated thrombosis (CAT). Treatment modalities were grouped as "none", "single", "targeted/hormonal", and "combination" to reduce model sparsity.

3.3. *Data pre-processing and exploratory data analysis*

RStudio was used to clean, recode, and analyse the data. The population characteristics were summarised using descriptive statistics. Categorical variables were reported as counts and percentages, while continuous variables were reported as means and standard deviations. Comparison between those with and without CAT were assessed using independent t-test for continuous variables and Fisher's exact test for categorical variables. There was no imputation of missing values because the study was conducted as a complete case analysis. Survival probabilities over time were estimated using Kaplan-Meier curves to accommodate censored observations. To determine if the survival distributions differed by CAT status, this study employed the log-rank test. This test evaluates the null hypothesis of no difference between the groups` survival experiences, with a significant p value denoting a statistically significant divergence in the curves (Kleinbaum & Klein 1996).

3.4. *Model development and statistical analysis*

This study applies two survival model which are the Cox proportional hazards (Cox PH) model and the log-logistic survival regression (LLSR) model to estimate overall survival in CAT patients. The Cox PH model was chosen for its semi-parametric flexibility and the interpretability of its hazard ratios under proportional hazards assumption. In contrast, the log-logistic (LLSR) model was chosen to accommodate potential non-monotonic hazard functions

and provide accurate survival probability estimates when hazards are not proportional. The Cox PH model is usually written in terms of the hazard function as shown in Eq. (1), where $h_0(t)$ is the baseline hazard at time t, $X_1, X_2, \dots, X_P$ is the covariates, and $\beta_1, \beta_2, \dots, \beta_P$ is the coefficients associated with the covariates. The hazard function for LLSR is shown in Eq. (2) and the survival function is shown in Eq. (3), where scale parameter $\lambda > 0$ and shape parameter $p > 0$ (Kleinbaum & Klein 1996).

$$h(t|x) = h_0(t) \cdot \exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right) \tag{1}$$

$$h(t) = \frac{\lambda p t^{p-1}}{1+\lambda t^p} \tag{2}$$

$$S(t) = \frac{1}{1+\lambda t^p} \tag{3}$$

Eqs. (2) and (3) represent the baseline Log-Logistic hazard and survival functions, denoted as $h_0(t)$ and $S_0(t)$ in the absence of covariates (where $x = 0$). While the Log-Logistic distribution is frequently used in an Accelerated Failure Time (AFT) framework, where covariates act as a multiplier on time, this study adopts a Parametric Proportional Hazards (PH) specification. In this PH formulation, the covariates enter the model multiplicatively to scale the baseline hazard as follows:

$$h(t|x) = h_0(t) \cdot \exp(\beta^T x) \tag{4}$$

where $x = \left(X_1, \dots, X_p\right)^T$ and $\beta = \left(\beta_1, \dots, \beta_p\right)^T$. For the log-logistic baseline, $S_0(t) = \{1 + \lambda t^p\}^{-1}$, which implies the baseline cumulative hazard,

$$H_0(t) = -\log S_0(t) = \log(1 + \lambda t^p)$$

Hence,

$$H(t|x) = \exp(\beta^T x) H_0(t), \qquad S(t|x) = \exp\{-H(t|x)\} = (1 + \lambda t^p)^{-\exp(\beta^T x)} \tag{5}$$

Under this Proportional Hazards (PH) model, $\exp\left(\beta_j\right)$ quantifies the Hazard Ratio (HR) associated with a one-unit increase in x_j, holding other covariates constant. A hazard ratio $\exp\left(\beta_j\right) > 1$ indicates an increased risk (hazard) of the event occurring, effectively shifting the survival curve $S(t|x)$ downward relative to the baseline.

To conclude, under the specified Proportional Hazards (PH) framework, the covariates x act multiplicatively on the baseline hazard function $h_0(t)$ This differs from the common Accelerated Failure Time (AFT) interpretation of the log-logistic model, where covariates would instead scale the time variable t or the scale parameter λ. By adopting the PH specification in Eqs. (4) and (5), the term $\exp\left(\beta_j\right)$ quantifies the hazard ratio associated with a one-unit increase in x_j, holding other covariates constant. A hazard ratio greater than 1 $\left(\beta_j > 0\right)$ indicates an increase in the conditional probability of the event occurring at time t, whereas a ratio less than 1 $\left(\beta_j < 0\right)$ suggests a protective effect.

The same covariates were included into both models, namely, age group, gender, BMI category, ECOG performance status, smoking status, cancer type, cancer stage, disease status, cancer treatment, and CAT status. Key assumptions were validated through model diagnostics.

The proportional hazards assumption of Cox PH model was assessed using Schoenfeld residuals. We looked at both individual and global tests, and the assumption was considered met if the *p*-values were more than 0.05. The "flexsurvreg" function was used to fit the LLSR model with a log-logistic distribution. Covariates effects on survival time were interpreted using time ratios.

3.5. *Model evaluation and selection*

Several evaluation metrics were used to evaluate model performance. Model fit and parsimony were evaluated using the Akaike Information Criterion (AIC). AIC can be defined as:

$$AIC = 2k - 2log(L) \tag{6}$$

where k is the number of model parameters and L is the maximized value of the likelihood function. A better model fit is indicated by lower AIC (Akaike 2003). Discriminative ability was assessed using Harrell`s concordance index (C-index). The C-index evaluates the model`s ability to correctly rank-order the survival times of pairs of individuals based on their predicted risk scores (Harrell *et al*. 1982). Additionally, each patient's survival probabilities were estimated 365 days after diagnosis. Patients were classified into binary classifications (high vs. low risk of mortality within a year) based on a threshold of 0.5, and predictive accuracy was then evaluated in terms of accuracy (Eq. (7)), sensitivity (Eq. (8)), specificity (Eq. (9)), and Area Under the Receiver Operating Characteristic curve (AUROC). The metrics were calculated as:

$$\text{Accuracy} = \frac{\text{True Positives+True Negatives}}{\text{(True Positives+True Negatives+False Positives+False Negatives)}} \tag{7}$$

$$\text{Sensitivity (True Positive Rate)} = \frac{\text{True Positives}}{\text{(True Positives+False Negatives)}} \tag{8}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{(True Negatives+False Positives)}} \tag{9}$$

All patients included in this study had complete follow-up information available up to 365 days after diagnosis. No individuals were lost to follow-up or censored prior to this time point, ensuring that the 365-day survival status was fully observed for all cases.

For classification-based evaluation, predicted survival probabilities at 365 days were dichotomised into high- and low-risk groups using a threshold of 0.5. This threshold corresponds to the standard decision boundary under equal misclassification costs and was selected to provide an interpretable and consistent benchmark for comparing model performance. Accuracy, sensitivity, and specificity were therefore reported as complementary, threshold-dependent measures. To mitigate reliance on any single threshold, threshold-free metrics—including Harrell's concordance index (C-index) and the Area Under the Receiver Operating Characteristic curve (AUROC) were also used to assess the overall discriminative ability.

4. Results

4.1. *Study population characteristics*

4.1.1. Descriptive statistics

Table 1 summarizes the characteristics of the 249 cancer patients. The cohort comprised both male and female patients, with comparable proportions across gender. Besides providing descriptive data, this table reports *p*-values for exploratory comparisons between the CAT and No CAT groups to highlight potential differences in clinical and laboratory characteristics at baseline. Accordingly, these results provide contextual insight to inform subsequent multivariable modelling, rather than to support confirmatory hypothesis testing or establish causal associations.

Table 1: Exploratory comparison of baseline characteristics of cancer patients by CAT status

Covariate	Overall *n* = 249	No CAT *n* = 230	CAT *n* = 19	*p*-value
Gender				>0.9[a]
Male	125 (50.2%)	115 (50%)	10 (52.6%)	
Female	124 (49.8%)	115 (50%)	9 (47.4%)	
Ethnicity				0.2[a]
Malay	137 (55.0%)	122 (53.0%)	15 (78.9%)	
Chinese	83 (33.3%)	80 (34.8%)	3 (15.8%)	
Indian	24 (9.6%)	23 (10.0%)	1 (5.3%)	
Others	5 (2.0%)	5 (2.2%)	0 (0.0%)	
Smoking status				0.6[a]
Non-smoker	158 (63.5%)	147 (63.9%)	11 (57.9%)	
Smoker	91 (36.5%)	83 (36.1%)	8 (42.1%)	
ECOG performance status				0.007[a]
<2	215 (86.3%)	203 (88.3%)	12 (63.2%)	
≥2	34 (13.7%)	27 (11.7%)	7 (36.8%)	
BMI category				0.042[a]
Underweight	25 (10.0%)	21 (9.1%)	4 (21.1%)	
Normal	130 (52.2%)	122 (53.0%)	8 (42.1%)	
Overweight	51 (20.5%)	50 (21.7%)	1 (5.3%)	
Obese	43 (17.3%)	37 (16.1%)	6 (31.6%)	
Diabetes				0.8[a]
No	168 (67.5%)	156 (67.8%)	12 (63.2%)	
Yes	81 (32.5%)	74 (32.2%)	7 (36.8%)	
Hypertension				0.3[a]
No	139 (55.8%)	126 (54.8%)	13 (68.4%)	
Yes	110 (44.2%)	104 (45.2%)	6 (31.6%)	
Cardiovascular				>0.9[a]
No	234 (94.0 %)	216 (93.9%)	18 (94.7%)	
Yes	15 (6.0%)	14 (6.1%)	1 (5.3%)	
Other comorbidities				0.2[a]
No	119 (47.8%)	113 (49.1%)	6 (31.6%)	
Yes	130 (52.2%)	117 (50.9%)	13 (68.4%)	
Cancer stage				0.025[a]
Stage I	22 (8.8%)	20 (8.7%)	2 (10.5%)	
Stage II	38 (15.3%)	38 (16.5%)	0 (0.0%)	
Stage III	78 (31.3%)	75 (32.6%)	3 (15.8%)	
Stage IV	111 (44.6%)	97 (42.2%)	14 (73.7%)	
Covid-19 history				0.4[a]
No	194 (77.9%)	181 (78.7%)	13 (68.4%)	
Yes	55 (22.1%)	49 (21.3%)	6 (31.6%)	
Age, years, mean (SD)	58 (14)	58 (13)	52 (17)	0.14[b]

Table 1 (Continued)				
Albumin, g/dL, mean (SD)	33.4 (6)	33.6 (6)	30.5 (4.9)	0.017[b]
Platelet, $\times 10^9$/L, mean (SD)	329 (112)	327 (110)	361 (134)	0.3[b]
Haemoglobin, g/dL, mean (SD)	11.68 (1.94)	11.71 (1.94)	11.25 (1.83)	0.3[b]
Total white cells, $\times 10^9$/L, mean (SD)	9.1 (4.5)	9.0 (4.6)	10.4 (4.1)	0.2[b]

n (%); Mean (SD)

[a]Fisher's exact test; [b]Welch Two Sample t-test

Based on overall cohort, majority of patients were Malay (55%), followed by Chinese (33.3%), Indian (9.6%), and a small percentage of other ethnics (2%). Most patients were non-smoker (63.5%) and 86.3% of the patients had ECOG performance status less than 2. Most patients had normal BMI (52.2%), 10% of the patients were underweight, 20.5% were overweight, and 17.3% were obese. Regarding comorbidities, 32.5% had diabetes, 44.2% had hypertension, 6% had cardiovascular disease, and 52.2% had other comorbidities. Most patients were diagnosed at stage IV cancer (44.6%), followed by stage III (31.3%), stage II (15.3%), and stage I (8.8%). A history of COVID-19 infection was reported in 22.1% of the cohort. The mean age was 58 years (SD = 14). The mean albumin level was 33.4 g/dL (SD = 6), while mean platelet count, hemoglobin, and total white cell count were 329 $\times 10^9$/L (SD = 112), 11.68 g/dL (SD = 1.94), and 9.1 $\times 10^9$/L (SD = 4.5), respectively. Percentages of CAT reported across cancer stages were calculated using the total number of patients within each cancer stage as the denominator, rather than the overall cohort.

Besides, Table 1 presents baseline characteristics of patients stratified by CAT status, with *p*-values reported to provide exploratory comparisons between groups and to highlight potential differences in clinical and laboratory characteristics at baseline. In this cohort, exploratory analysis identified several baseline characteristics that potentially differ between patients with and without Cancer-Associated Thrombosis (CAT). Specifically, patients with CAT were more likely to present with an ECOG performance status ≥ 2 ($p = 0.007$), advanced Stage IV cancer ($p = 0.025$) and lower mean albumin levels ($p = 0.017$). Additionally, a higher proportion of obese patients was observed in the CAT group compared to those without CAT ($p = 0.042$).

While these variables showed statistical significance in our univariate exploratory comparisons, these results must be interpreted with caution. Because multiple comparisons were performed without adjustment for multiplicity, these findings are intended to be hypothesis-generating rather than confirmatory. These characteristics provide important context for the study and suggest potential factors that may influence thrombotic risk; however, their independent associations should be validated through further multivariable modeling and confirmatory studies.

Figure 1 illustrates the prevalence of thrombosis (n = 19) across different cancer types. In our study population, gastrointestinal cancer had the largest number of patients, but only 5.7% had CAT. Notable CAT proportions were also seen in breast and gynaecology, head and neck, and miscellaneous cancer, accounting for 6.2%, 6.9%, and 19.4% of patients, respectively. The genitourinary cancer, on the other hand, did not have any cases of thrombosis.

4.1.2. Kaplan-Meier survival analysis

Figure 2 shows the Kaplan-Meier survival curves to compare overall survival of patients with and without CAT. It showed that patients that were free from CAT had a higher survival probability over time, indicating they were less likely to experience the event (death). While patients with CAT showed a higher risk of mortality.

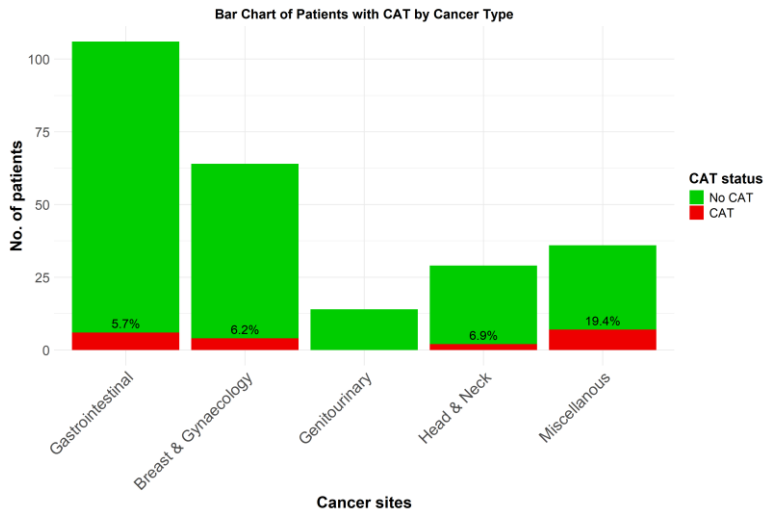

Figure 1: Bar chart of patients with CAT by cancer types

Table 2 shows that the No-CAT group was larger than the CAT group, with 230 patients who did not develop CAT compared to 19 patients who did. Patients with CAT had an observed median survival time of 391 days, while those without CAT had a median survival time of 432 days. It should be noted that the No-CAT group was considerably larger than the CAT group, which may have affected the stability of the median estimates. However, the log-rank test yielded a *p*-value of 0.01, indicating a statistically significant difference in the overall survival distributions between patients with and without CAT. These results suggest that survival among cancer patients is reduced when CAT is present

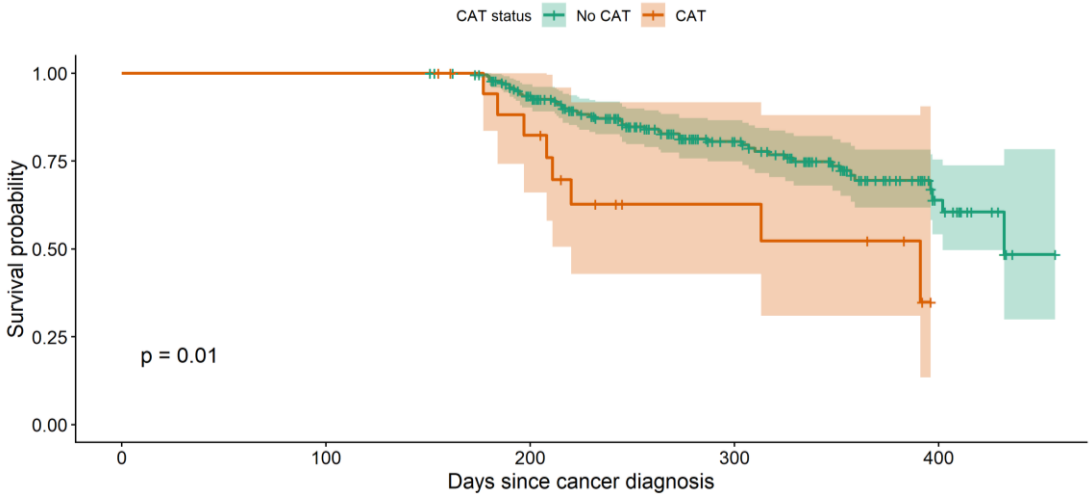

Figure 2: Kaplan-Meier survival curve by CAT status

Table 2: Log-rank test table

Group	No. of patients	Median of survival	Log-rank *p*-value
No CAT	230	432	0.01
CAT	19	391	

4.2. *Model development and performance*

In the multivariable survival models, likelihood ratio (LR) *p*-values are reported to assess the global significance of each covariate, reflecting whether the variable contributes to the model. In contrast, Wald-based hazard ratios (HRs) or time ratios (TRs), together with their 95% confidence intervals and Wald *p*-values, correspond to individual category-level contrasts relative to the reference group. Accordingly, while a covariate may demonstrate global significance based on the LR test, individual contrasts with confidence intervals crossing 1 are interpreted cautiously and are not described as statistically significant.

4.2.1. Cox proportional hazards model

Table 3 summarizes the results from the multivariable Cox PH model. ECOG performance status (p = <0.001), cancer stage (p = <0.001), cancer treatment group (p = 0.047), and disease status (p = 0.027) were significantly associated with overall survival based on the likelihood ratio test. Specifically, patients with poor ECOG performance status (≥2) had a higher risk of death compared to those with better ECOG performance (HR = 3.04; 95% CI = 1.49–6.21). Cancer stage demonstrated a statistically significant global association with overall survival based on the likelihood ratio test (p < 0.001). At the individual contrast level, patients with Stage IV cancer exhibited a higher hazard of death compared with Stage I (HR = 4.01, 95% CI: 1.17–13.66), whereas other stage-specific contrasts showed elevated hazards but did not reach statistical significance, as their confidence intervals included 1.

Table 3: Cox PH model summary table

Covariate	HR (95% CI)	Wald *p*-value	Likelihood ratio test *p*-value
ECOG ≥ 2 vs <2	3.04 (1.49 – 6.21)	0.002 **	<0.001 ***
Normal vs Underweight	0.37 (0.15 – 0.91)	0.030 *	0.196
Overweight vs Underweight	0.40 (0.14 – 1.11)	0.080	
Obese vs Underweight	0.41 (0.14 – 1.24)	0.115	
Breast & Gynecology vs Gastrointestinal	0.73 (0.30 – 1.82)	0.504	0.063
Genitourinary vs Gastrointestinal	0.20 (0.03 – 1.52)	0.119	
Head & Neck vs Gastrointestinal	0.37 (0.13 – 1.07)	0.066	
Miscellaneous vs Gastrointestinal	1.11 (0.51 – 2.45)	0.790	
Stage II vs Stage I	0.19 (0.02 – 1.84)	0.151	<0.001 ***
Stage III vs Stage I	1.74 (0.46 – 6.61)	0.417	
Stage IV vs Stage I	4.01 (1.17 – 13.66)	0.027 *	
Single treatment vs No treatment	1.19 (0.42 – 3.36)	0.737	0.047 *
Targeted/ Hormonal vs No treatment	0.28 (0.03 – 2.65)	0.264	
Combination vs No treatment	0.50 (0.19 – 1.31)	0.159	
Recurrent vs Active	2.80 (0.82 – 9.60)	0.101	0.027 *
CAT vs No CAT	1.03 (0.43 – 2.49)	0.948	0.948

Both cancer treatment group and disease status were also statistically significant based on likelihood ratio test, suggesting they contributed meaningfully to model fit. Compared to patients who received no treatment, those who received a single treatment such as surgery, chemotherapy, or radiation had a hazard ratio of 1.19 (95% CI = 0.42–3.36). Those who received targeted or hormonal therapy had a hazard ratio of 0.28 (95% CI = 0.03–2.65), and those who received combination therapy had a hazard ratio of 0.50 (95% CI = 0.19–1.31). For

the disease status, patients with recurrent disease had a higher mortality risk compared to active disease status (HR = 2.80; 95% CI = 0.82–9.60).

Although the likelihood ratio test suggested a borderline global association between BMI and mortality (p = 0.196), the individual contrast comparing normal BMI with underweight indicated a lower hazard of death (HR = 0.37, 95% CI: 0.15–0.91), which should be interpreted cautiously. Cancer type showed borderline significance based on the likelihood ratio test (p = 0.063). Although this did not reach statistical significance in general, some cancer sites showed notable patterns. The risk of mortality was lower for patients with head and neck cancers compared to those with gastrointestinal cancers (HR = 0.37; 95% CI: 0.13–1.07), and the p-value based on the Wald test was close to significance (p = 0.066). Similarly, genitourinary cancers were associated with a lower mortality risk (HR = 0.20; 95% CI: 0.03–1.52), although not statistically significant (p = 0.119). The proportional hazards assumption was tested using Schoenfeld residuals and there was no violation (global p-value = 0.52).

4.2.2. Log-logistic survival regression model

Table 4 presents the results summary for LLSR model. The LLSR model identified ECOG performance status (p = 0.001), cancer stage (p = <0.001), and disease status (p = 0.03) as significant variables. Patients with ECOG performance more than or equal to two had significantly shorter survival times compared to those with ECOG performance less than two (TR = 0.71; 95% CI = 0.59–0.87). Cancer stage was globally associated with survival time based on the likelihood ratio test (p < 0.001). At the individual contrast level, Stage III (Stage III TR = 0.80; 95% CI: 0.57–1.12) and Stage IV (Stage IV TR = 0.68; 95% CI: 0.49–0.93) patients exhibited shorter survival times compared with Stage I, as reflected by time ratios less than 1, however, several contrasts demonstrated wide confidence intervals, and results with confidence intervals crossing 1 were interpreted as not statistically significant. Recurrent patients were associated with shorter survival times compared to active patients (TR = 0.66; 95% CI = 0.46–0.94).

Table 4: LLSR model summary table

Covariate	TR (95% CI)	Wald p-value	Likelihood ratio test p-value
ECOG ≥ 2 vs <2	0.71 (0.59 – 0.87)	0.001 **	0.001 **
Breast & Gynecology vs Gastrointestinal	1.10 (0.87 – 1.40)	0.430	0.062
Genitourinary vs Gastrointestinal	1.67 (0.98 – 2.82)	0.057	
Head & Neck vs Gastrointestinal	1.29 (0.96 – 1.73)	0.090	
Miscellaneous vs Gastrointestinal	0.98 (0.79 – 1.21)	0.822	
Stage II vs Stage I	1.50 (0.87 – 2.58)	0.148	<0.001 ***
Stage III vs Stage I	0.80 (0.57 – 1.12)	0.186	
Stage IV vs Stage I	0.68 (0.49 – 0.93)	0.016 *	
Recurrent vs Active	0.66 (0.46 – 0.94)	0.021 *	0.030 *
Single treatment vs No treatment	0.88 (0.66 – 1.19)	0.418	0.075
Targeted/ Hormonal vs No treatment	1.30 (0.64 – 2.63)	0.463	
Combination vs No treatment	1.12 (0.85 – 1.48)	0.432	
CAT vs No CAT	0.98 (0.76 – 1.27)	0.897	0.897

Cancer type and treatment showed a borderline significance, p = 0.062, and p = 0.075, respectively. Although individual comparisons were not significant, genitourinary cancers patients showed a longer survival times compared to gastrointestinal cancers patients (TR =

1.67; 95% CI = 0.98–2.82; p = 0.057). For cancer treatment, patients who received targeted therapy or hormonal therapy had slightly longer survival times than those who received no treatment (TR = 1.30; 95% CI = 0.64–2.63; p = 0.463), though this was not statistically significant. Individual HRs and TRs whose 95% confidence intervals included 1 were interpreted as inconclusive and were not described as statistically significant, even when the corresponding covariate demonstrated global significance based on the likelihood ratio test.

4.3. *Model evaluation*

Model evaluation was conducted across three distinct dimensions: relative model fit, survival discrimination, and binary classification performance at 365 days. Relative model fit was assessed using the Akaike Information Criterion (AIC), where lower values indicate a better trade-off between model complexity and goodness-of-fit. Survival discrimination was evaluated using Harrell's concordance index (C-index), which measures the model's ability to correctly rank patients according to their survival risk over time. Classification performance at 365 days post-diagnosis was assessed separately using accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUROC), based on dichotomized survival status at one year.

In both survival models, likelihood ratio test p-values were used to assess the global significance of each covariate, while Wald-based hazard ratios (Cox PH) and time ratios (LLSR), together with their 95% confidence intervals, correspond to individual category-level contrasts.

Based on AIC, the Cox PH model demonstrated superior relative model fit compared with the LLSR model (534.67 vs. 839.93). In terms of survival discrimination, both models exhibited comparable performance as measured by Harrell's C-index, with values of 0.785 for the Cox PH model and 0.788 for the LLSR model.

Classification performance at 365 days showed modest differences between models. Using a probability threshold of 0.5, the Cox PH model achieved slightly higher accuracy (79.92%) and specificity (87.76%), whereas the LLSR model demonstrated marginally higher sensitivity (60.38%) and AUROC (80.46%). These findings indicate that while the Cox PH model provided better overall fit, both models showed similar ability to discriminate survival risk and classify one-year mortality status.

In the Cox PH model, covariate effects are expressed as hazard ratios, reflecting relative differences in instantaneous risk of death, whereas in the LLSR model, effects are expressed as time ratios, which quantify acceleration or deceleration of survival time. Accordingly, hazard and time ratios are interpreted within their respective modelling frameworks and are not directly comparable.

Table 5: Comparison of model performance between Cox PH and LLSR models

Metric	Cox PH Model	LLSR Model
AIC	534.6717	839.9262
Harrell's C-index	0.7854	0.7882
Accuracy	79.92%	79.52%
Sensitivity	50.94%	60.38%
Specificity	87.76%	84.69%
AUROC	80.32%	80.46%

5. Discussion

This study compared the Cox proportional hazards (PH) and log-logistic survival regression (LLSR) models in predicting all-cause mortality among cancer patients. Several clinical factors, including ECOG performance status, cancer stage, cancer treatment group, and disease status, showed global statistical significance when assessed by the likelihood ratio test in both models. Based on individual parameter estimates, patients with poor ECOG performance (≥ 2) and advanced cancer stage (Stage III and IV) demonstrated a significantly higher hazard ratio in the Cox PH model, while similar trends regarding shorter survival times were observed in the LLSR model.

The models were evaluated using three distinct approaches: relative model fit, survival discrimination, and 365-day binary classification performance. Based on the Akaike Information Criterion (AIC), the Cox PH model demonstrated a better relative model fit compared to the LLSR model. Harrell's C-index, which measures discriminative ability regarding survival risk ranking, showed similar performance between the two models. When assessing binary classification performance at 365 days, the Cox PH model displayed higher accuracy and specificity, whereas the LLSR model yielded slightly higher sensitivity and AUROC.

According to the Cox PH model, patients with poor ECOG performance (≥ 2) had a significantly higher mortality risk (Hazard Ratio [HR] > 1, 95% CI > 1), and those with advanced cancer stage (Stage IV) also showed significantly higher mortality risk compared to Stage I. The LLSR model supported these findings, associating ECOG (≥ 2) and advanced cancer stage with shorter survival time ratios. While cancer treatment groups were significant globally in the Cox model, individual therapy comparisons (e.g., single treatments vs. combination) frequently resulted in confidence intervals crossing 1, suggesting these specific contrasts require validation in larger cohorts. Similarly, while the global likelihood ratio test for BMI was not significant ($p = 0.196$), individual Wald test estimates indicated that patients with normal BMI had a lower mortality risk compared to underweight patients.

In both the Cox PH and LLSR models, CAT status did not reach statistical significance. However, the observed direction of the effect consistently indicated that patients with CAT may face a higher mortality risk and reduced survival compared to those without CAT. Although this difference was not statistically significant in our cohort, the observed pattern aligns with existing literature, such as findings by Mulder *et al*. (2021) and Lee *et al*. (2021), which similarly demonstrated that cancer patients with thrombosis exhibit a higher probability of death. This lack of statistical significance in our study may be attributed to a low event rate, the limited sample size, or potential confounding by other unmeasured clinical factors, suggesting that this association warrants further investigation in larger, multicenter cohorts

6. Conclusion

In cancer patients with CAT, the Cox PH and LLSR models are both useful for determining important prognostic variables and estimating survival. The two most important variables influencing survival in both models were ECOG performance status and cancer stage. Despite not being statistically significant, CAT status demonstrated a consistent direction of effect, which align with the existing studies on its clinical impact. The model evaluation performance showed that Cox PH model had slightly higher classification performance at one year and better model fit. This means that the Cox PH model was more accurate in predicting which patients survived and which patients experienced the event within 1 year of follow-up. The small sample size and absence of time-dependent factors are the two limitations of this study. The small

overall sample size, particularly the smaller size of patients with CAT compared to patients without CAT forms an imbalance that might affect the accuracy and stability of the survival predictions in CAT patients. The absence of time-dependent features also might influence hazard assumptions over time, regardless of being strengthened by real-world comprehensive data and a comparison of two modelling approaches. Prognostic modelling in this population may be further improved by future research that includes larger cohorts and time-dependent factors.

Acknowledgments

We would like to thank the Universiti Teknologi MARA, UiTM for supporting this research study under FRGS Research Grant, 600-RMC/FRGS 5/3 (068/2024). Besides, this study was approved by the Research Ethics Committee of Universiti Kebangsaan Malaysia (FF-2022-099) and the Medical Research and Ethics Committee (MREC) of the Malaysian Ministry of Health (MOH) (NMRR ID-22-01438-MW4 (IIR)), and it is in accordance with the Helsinki Declaration (IV adaptation).

References

Abbas L., Muhamad Nor I., Ismail F., Palaniappan S., Chong G.Y., Wan Puteh S.E., Muhamad Jamil S.A. & Tumian N.R. 2025. Factors associated with thrombosis among solid organ cancer patients in Kuala Lumpur, Malaysia. *Thrombosis Journal* **23**(1): 25.

Akaike H. 2003. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6): 716–723.

Ay C., Pabinger I. & Cohen A.T. 2017. Cancer-associated venous thromboembolism: Burden, mechanisms, and management. *Thrombosis and Haemostasis* **117**(02): 219–230.

Bardo M., Huber C., Benda N., Brugger J., Fellinger T., Galaune V., ... & Friede T. 2024. Methods for non-proportional hazards in clinical trials: A systematic review. *Statistical Methods in Medical Research* **33**(6): 1069–1092.

Borges P. 2021. Estimating the turning point of the log-logistic hazard function in the presence of long-term survivors with an application for uterine cervical cancer data. *Journal of Applied Statistics* **48**(2): 203–213.

Carrier M., Abou-Nassar K., Mallick R., Tagalakis V., Shivakumar S., Schattner A., ... & Wells P.S. 2019. Apixaban to prevent venous thromboembolism in patients with cancer. *New England Journal of Medicine* **380**(8): 711–719.

Crobach M.J.T., Anijs R.J.S., Brækkan S.K., Severinsen M.T., Hammerstrøm J., Skille H., ... & Cannegieter S.C. 2023. Survival after cancer-related venous thrombosis: The Scandinavian Thrombosis and Cancer Study. *Blood Advances* **7**(15): 4072–4079.

Deo S.V., Deo V. & Sundaram V. 2021. Survival analysis—part 2: Cox proportional hazards model. *Indian Journal of Thoracic and Cardiovascular Surgery* **37**(2): 229–233.

Donnellan E. & Khorana A.A. 2017. Cancer and venous thromboembolic disease: A review. *The Oncologist* **22**(2): 199–207.

Fallon M., Giusti R., Aielli F., Hoskin P., Rolke R., Sharma M., Ripamonti C.I. & ESMO Guidelines Committee. 2018. Management of cancer pain in adult patients: ESMO Clinical Practice Guidelines. *Annals of Oncology* **29** (Suppl. 4): iv166–iv191.

Fernandes C.J., Morinaga L.T.K., Alves Jr J.L., Castro M.A., Calderaro D., Jardim C.V.P. & Souza R. 2019. Cancer-associated thrombosis: The when, how and why. *European Respiratory Review* **28**(151): 180119.

Harrell F.E., Califf R.M., Pryor D.B., Lee K.L. & Rosati R.A. 1982. Evaluating the yield of medical tests. *JAMA* **247**(18): 2543–2546.

Hashemian A.H., Garshasbi M., Pourhoseingholi M.A. & Eskandari S. 2017. A comparative study of cox regression vs. log-logistic regression (with and without its frailty) in estimating survival time of patients with colorectal cancer. *Journal of Medical and Biomedical Sciences* **6**(1): 35–43.

Khan U.T., Walker A.J., Baig S., Card T.R., Kirwan C.C. & Grainge M.J. 2017. Venous thromboembolism and mortality in breast cancer: Cohort study with systematic review and meta-analysis. *BMC Cancer* **17**(1): 747.

Kleinbaum D.G. & Klein M. 1996. *Survival analysis a self-learning text*. New York, NY: Springer.

Kumar M., Sonker P.K., Saroj A., Jain A., Bhattacharjee A. & Saroj R.K. 2020. Parametric survival analysis using R: Illustration with lung cancer data. *Cancer Reports* **3**(4): e1210.

Lee L.H., Nagarajan C., Tan C.W. & Ng H.J. 2021. Epidemiology of cancer-associated thrombosis in Asia: A systematic review. *Frontiers in Cardiovascular Medicine* **8**: 669288.

McLernon D.J., Giardiello D., Van Calster B., Wynants L., van Geloven N., van Smeden M., ... & topic groups 6 and 8 of the STRATOS Initiative. 2023. Assessing performance and clinical usefulness in prediction models with survival outcomes: Practical guidance for Cox proportional hazards models. *Annals of Internal Medicine* **176**(1): 105–114.

Mulder F.I., Horvàth-Puhó E., van Es N., Van Laarhoven H.W.M., Pedersen L., Moik F., ... & Sørensen H.T. 2021. Venous thromboembolism in cancer patients: A population-based cohort study. *Blood, The Journal of the American Society of Hematology* **137**(14): 1959–1969.

Peng Q., Zhu J., Zhang Y. & Jing Y. 2024. Blood hypercoagulability and thrombosis mechanisms in cancer patients-A brief review. *Heliyon* **10**(19): e38831.

Wang Z., Lee J.W., Chakraborty T., Ning Y., Liu M., Xie F., Ong M.E.H. & Liu N. 2024. Survival modeling using deep learning, machine learning and statistical methods: A comparative analysis for predicting mortality after hospital admission. *arXiv preprint arXiv:2403.06999*.

Zhang Z., Reinikainen J., Adeleke K.A., Pieterse M.E. & Groothuis-Oudshoorn C.G. 2018. Time-varying covariates and coefficients in Cox regression models. *Annals of Translational Medicine* **6**(7): 121.

Center of Mathematical Sciences
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450, Shah Alam
Selangor, MALAYSIA
E-mail: adlinahajazi@gmail.com, afiqahjamil@uitm.edu.my, shahida@tmsk.uitm.edu.my*

Data Driven and Modelling (DDM) Research Initiative Group (RIG)
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam
Selangor, MALAYSIA
*Email: afiqahjamil@uitm.edu.my**

Unit of Internal Medicine
Hospital Tangkak
Ministry of Health, MALAYSIA
E-mail: lailatulema@googlemail.com

Department of Medicine
Faculty of Medicine
Universiti Kebangsaan Malaysia
Jalan Yaacob Latif, 56000
Kuala Lumpur, MALAYSIA

Hospital Canselor Tuanku Muhriz
Jalan Yaacob Latif, 56000
Kuala Lumpur, MALAYSIA
E-mail: rafeah@hctm.ukm.edu.my

Received: 24 July 2025
Accepted: 7 April 2026

*Corresponding author