Afrika Statistika

Afrika Statistika
Vol. 18 (1), 2023, pages 3379 - 3398.
DOI: http://dx.doi.org/10.16929/as/2023.3379.310

ISSN 2316-090X

Model selection criteria for survival data based on Kullback's divergence: A systematic and critical review

Houénafa Clarisse Dete [1,*], **Marcel Senou** [2], **Gneyou Essona Kossi** [3] **and Romain Glèlè Kakaï** [1]

[1] Laboratoire de Biomathématiques et d'Estimations Forestières, Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 04 BP 1525 Cotonou, Benin.
[2] Faculté des Sciences Agronomiques (FSA), Université d'Abomey-Calavi, 01BP: 526 Cotonou (Bénin)
[3] Department of Mathematics, Université de Lomé, BP 1515 Lomé, Togo.

Received on July 7, 2023; Accepted on September 25, 2023; Published on Octobre 29, 2023

Abstract We did a literature review to summarize the trends in the model selection criteria derived from Kullback's divergence in survival analysis. Furthermore, we conducted comprehensive discussions on these criteria to enhance the users' understanding. Therefore, 4628 original papers on model selection criteria in survival analysis are identified via keyword searching using Pubmed, Web of Science, and Scopus search engines. Subsequently, 304 studies were fully analyzed, excluding those that did not utilize criteria based on Kullback's divergence for model selection. The most commonly reported model selection criteria were the AIC and the AIC_c. Surprisingly, none of the selected papers discussed of the KIC family model selection criteria.

Key words: Survival analysis, model selection criteria, Kullback's divergence, performance.
AMS 2010 Mathematics Subject Classification Objects : 62N01; 62F07; 62G20.

[*] Houénafa Clarisse Dete: clarissedete@gmail.com
Marcel Senou: m_senou2002@yahoo.fr
Gneyou Essona Kossi: kossi_gneyou@yahoo.fr
Romain Glèlè Kakaï : glele.romain@gmail.com

Résumé. (Abstract in French) Nous avons procédé à une revue systhématique pour résumer et pour discuter des critères de sélection des modèles dérivés de la divergence de Kullback utilisés en analyse de survie afin d'améliorer la compréhension des utilisateurs. Pour y arriver, 4628 articles originaux sur les critères de sélection des modèles en analyse de survie ont été identifiés à l'aide d'une recherche par mots-clés sur Pubmed, Web of Science et Scopus. Ensuite, 304 études ont été analysées, en excluant celles qui n'employaient pas de critères basés sur la divergence de Kullback. Les critères de sélection des modèles les plus fréquemment rapportés étaient le AIC et le AIC_c. Mais force est de constater qu'aucun des articles sélectionnés ne traite des critères de sélection des modèles de la famille des KIC.

Presentation of authors.

Houénafa Clarisse Dete, M.Sc., is preparing a Ph.D. thesis under the third and the fourth authors at: Laboratoire de Biomathématiques et d'Estimations Forestières, Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 04, BP 1525, Cotonou, Benin

Marcel Senou, Ph.D., is a Full Professor is of Quantitative Genetics at: Faculté des Sciences Agronomiques (FSA), Université d'Abomey-Calavi, 01BP: 526 Cotonou (Bénin).

Kossi Essona Gneyou, Ph.D., is Full Professor of Statistics and Mathematics at: Department of Mathematics, Université de Lomé, BP 1515 Lomé, Togo.

Romain Glèlè Kakaï, Ph.D., is Full Professor of Biostatistics and Forest estimations at: Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 04, BP 1525, Cotonou, Benin. He is member of the: Laboratoire de Biomathématiques et d'Estimations Forestières.

1. Introduction

Survival analysis is a subfield of statistics and refers to a set of primary statistical methods to analyze the span of time until a precise event occurs.(Schober and Vetter, 2018). It is a time-to-event analysis because it helps to respond to the issue of whether or not the occurrence occurred (Dichotomous outcome) and the time of its occurrence (continuous outcome) (Schober and Vetter, 2018). To analyze survival data, extensive statistical models are developed. The most widely used regression model is the Cox Proportional Hazards (PH) model (Cox, 1972) followed by Parametric Proportional Hazard models(Lu et al., 2015). The accelerated failure time (AFT) model, often called the log location scale (Lawless, 2011), offers an alternative to the PH model (Wei, 1992).

In the statistical modeling of survival data, choosing the best model to fit the underlying data within a collection of candidates is a crucial question (Kim

et al., 2014). A biased model may significantly impact scientific interpretations as well as model predictions. For this purpose, the appropriate model is chosen using model selection criteria. The Akaike Information Criterion (AIC) was the first widely accepted criteria (Akaike, 1973, 1974). But Hurvich and Tsai (1989) suggested using the corrected AIC (AIC_c) criterion for small samples. Additionally, an improved AIC was proposed (AIC_{sur}) along the same lines as AIC_c (Liang and Zou, 2008). AIC, AIC_c, and AIC_{sur} criteria have been developed using Kullback's Directed Divergence, which measures the divergence betwixt dissimilarity model and true model (Kullback, 1968). Other criteria derived from Kullback Symmetric Divergence (KSD) were developed, including Kullback Information Criterion (KIC), corrected KIC (KIC_c), and Improved KIC (KIC_{sur}) (Cavanaugh and Noe, 1999; Cavanaugh, 2004; Kim and Cavanaugh, 2005).

Even though many studies have been conducted to assess the performance of the numerous model selection criteria in survival analysis and offer new solutions, a systematic review of criteria in survival analysis is uncommon. However, survival analysis has emerged as one of the most common data analysis techniques in various fields, including medicine, criminology, marketing, astronomy, epidemiology, and environmental health, in the last four decades (Shaik et al., 2015). Users can only properly use the model selection criteria as helpful tools in identifying the "best" potential model provided they are aware of the strengths and limits of the criteria. Therefore, it is crucial to provide a thorough overview of the various model selection criteria to aid in selecting statistically sound models. In this study, we aim to (i) introduce the various criteria derived from Kullback's divergence used in survival analysis, (ii) analyze their strengths and weaknesses according to past research studies, and (iii) identify research gaps in the model selection criteria in survival analysis in addition to putting forth research perspectives.

The remainder of the work is structured as follows: Section 2 describes the methodology, Section 3 contains our results and discussions, and Section 4 addresses the gaps and perspectives. A brief conclusion, in Section 5, ends the paper.

2. Methods

Electronic scientific databases, like Pubmed, Web of Science, and Scopus, were consulted to search for model selection criteria in survival analysis. We used a combination of various search terms for model selection criteria and survival analysis. The following keywords were used: "Akaike Information Criterion" OR "corrected Akaike Information Criterion" OR "improved Akaike Information Criterion" OR "Kullback Information Criterion" OR "corrected Kullback Information Criterion" OR "improved Kullback Information Criterion" OR" Takeuchi Information Criterion" OR " conditional Akaike Information Criterion" OR "marginal Akaike Information Criterion" OR "marginal Akaike Information Criterion" OR" model selection" OR" model selection criteria" AND "Survival analysis" OR "Survival data" OR "Censored data" OR "time-to-event analysis". The systematic review includes 304 papers from a preliminary selection of 4628. The papers selected in the

systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The main inclusion criterion used was the scope of papers (papers focused on model selection criteria in survival analysis). All studies of survival data that did not use information criteria based on KDD and KSD for model selection were excluded. Non-research papers, commentaries, and non-English language papers were also excluded. Papers dealing with criteria, such as Conceptual predictive, Bayesian Information Criterion, R^2, and Hannan–Quinn Information Criterion, were excluded too (Figure 1, page 3383).

The following information was retrieved for each of the 304 papers selected: the authors' names, the year of publication, the country where the study was done, the model selection criteria used, the models used, the areas of use, the distribution function, the censoring mechanism, the number of parameters and the sample size. These data enable the description of the criteria used and the evaluation of the trend using these model selection criteria in survival analysis over the past years. The Model selection criteria, distribution function, fields of use, and models used were presented using bar plots. The number of papers categorized based on the censoring mechanism or truncation was reported. Furthermore, the ratio of sample size to the number of model parameters was calculated and depicted using a bar plot, illustrating its relationship with the model selection criterion. Lastly, an exact Fisher test was performed to determine any potential association between the distribution and the model used.

3. Results and discussion

3.1. Paper characteristics

The majority of the 304 papers reviewed (59.50 %) were from Asia, particularly China (87) and Iran (36) (Figure 2(a), page 3384). However, in Europe, we observed a higher proportion of countries where studies were conducted compared to other continents. Specifically, studies were conducted in 29.54 % (13 countries out of 44) of the countries in Europe, with Belgium (7 studies), Germany(6 studies), and the United Kingdom (6 studies)standing out as the countries with the highest number of studies. Our sampled studies included 32 that focused on African countries, with the vast majority being conducted in Ethiopia. The chosen studies covered 5 out of 35 countries in the American, with the United States (31 studies) and Canada (12 studies) carrying out the majority of the research (Figure 2(b), page 3384).

The papers considered for the study demonstrate that the use of survival analysis has increased exponentially over time (Figure 3, page 3384). Most of the papers (53.16 %) were in the fields of medicine, followed by Mathematics and Statistics (16.95 %), Public Health, Environmental and Occupational Health (5.75 %), Engineering and Technology (4.88 %), and Experimental Medicine (4.02 %) (Figure 4, page 3385). Some of the mathematical and statistical papers were co-authored by health researchers. These results are consistent with Shaik et al. (2015) and Ra-

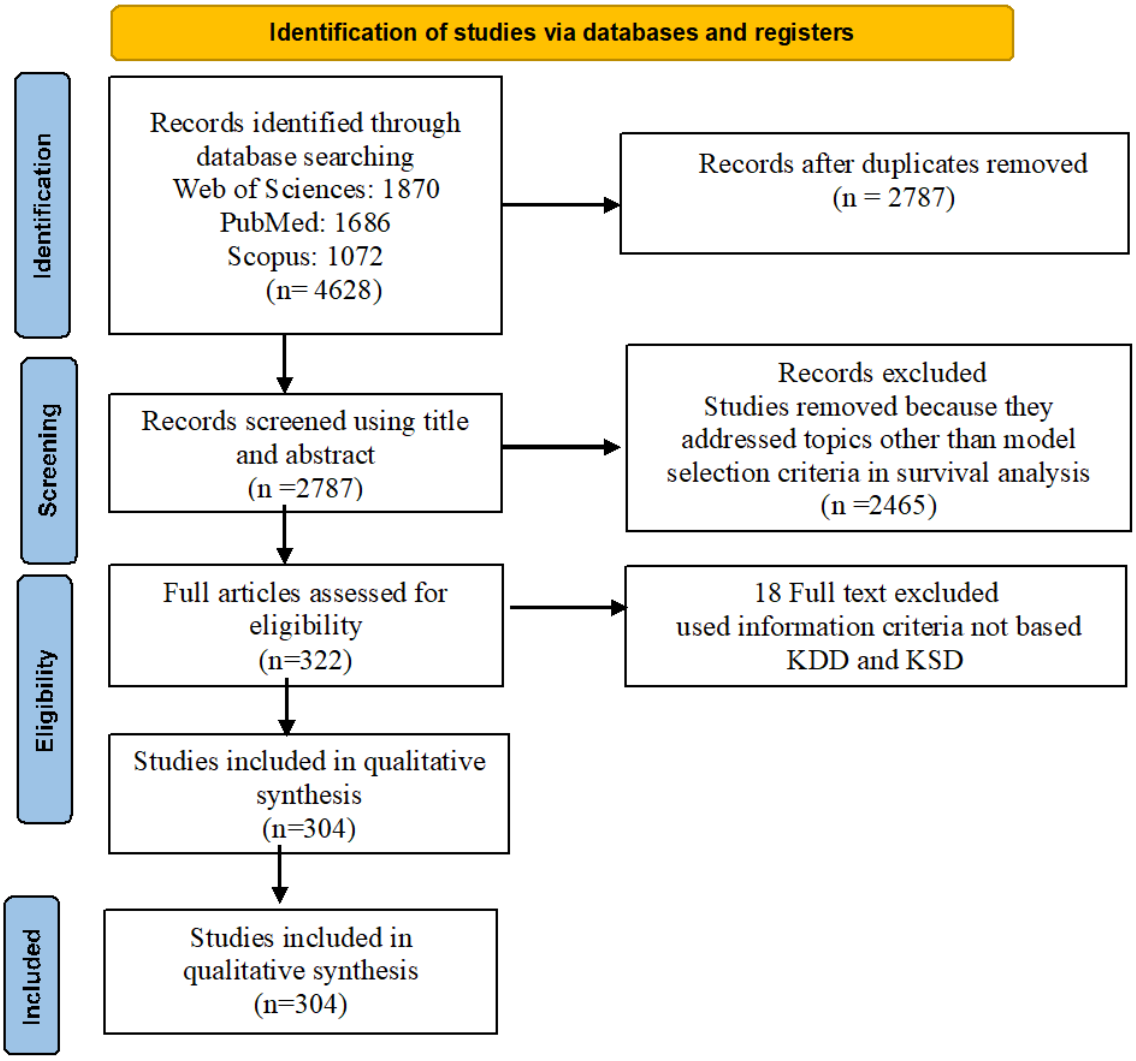

Figure 1. PRISMA flow diagram of the selection procedure.

masamy and Kaliannan (2021) as they stated that survival analysis has gained significant popularity as a favored method for analyzing survival time across various fields, including medicine, epidemiology, environmental health, and criminology marketing, and astronomy over the last four decades.

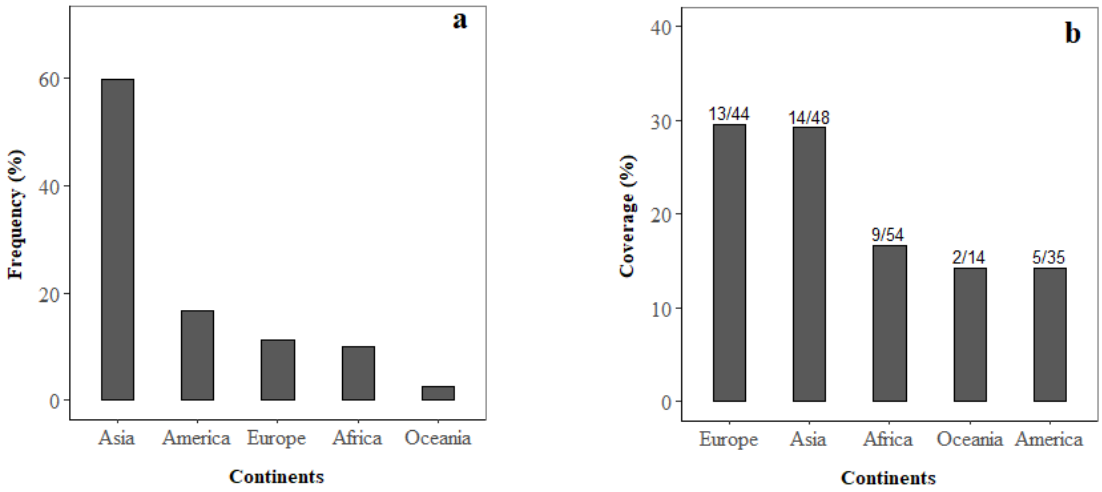

Figure 2. Distribution of research throughout continents (a) and coverage of countries throughout continents (b)

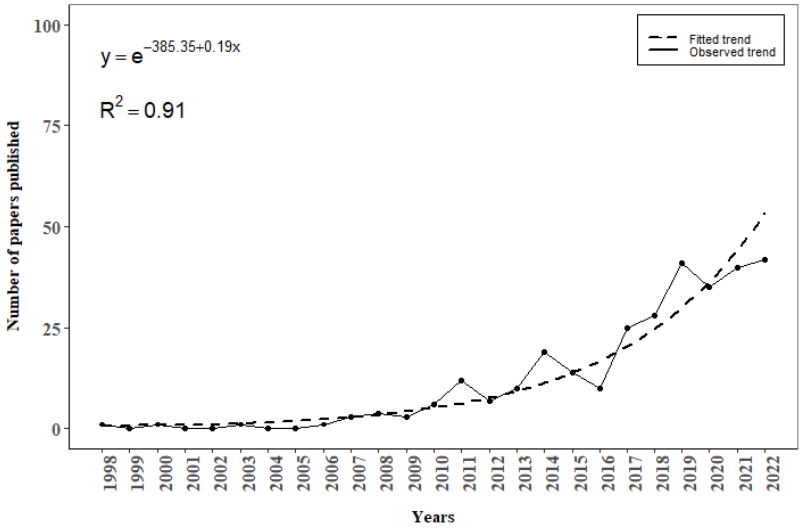

Figure 3. The trend of usage of survival analysis over the last two decades (from 1998 to 2022).

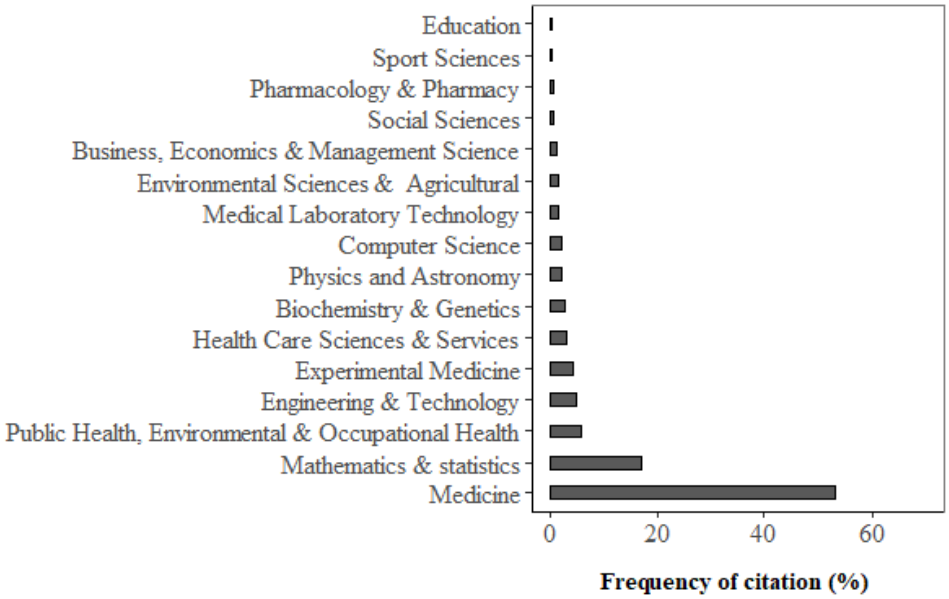

Figure 4. Research fields of selected papers.

3.2. Overview on model selection criteria based on Kullback's divergence

Let Y represent the observed data vector, θ_0 represent the unknown true parameter vector, and θ_k represent the parameter vector of the candidate model. The generating and candidate densities for data are represented by $h(Y|\theta_0)$ and $h(Y|\theta_k)$, respectively. Let θ_k be an estimate vector obtained by maximizing the likelihood function $h(Y|\theta_k)$ over Θ_k, and $h(Y|\hat{\theta}_k)$ signifies the associated fitted model. Assume our goal is to find the ajusted model $h(Y|\hat{\theta}_k); k \in \{k_1, k_2, \cdots, k_L\}$, which constitute the "best" approximation to $h(Y|\theta_0)$ amid a collection of families $\{\mathscr{H}(k_1), \mathscr{H}(k_2), \cdots, \mathscr{H}(k_L)\}$.

As a result, our model selection problem can be seen as a dimension determination problem. To identify which of the fitted models $h(Y|\theta_1), h(Y|\theta_2), \cdots, h(Y|\theta_L)$ most closely approaches $h(Y|\theta_0)$, we need a measure that gives an appropriate representation of the difference between the genuine model $h(Y|\theta_0)$ and an approximate model $h(Y|\theta_k)$. Kullback's Directed and Symmetric Divergences both achieve this objective (Cavanaugh and Neath, 2019).

3.2.1. **Criteria derived from Kullback's Directed Divergence**

The Kullback's Directed Divergence betwixt $h(Y|\theta_0)$ and $h(Y|\theta_k)$ with respect to $h(Y|\theta_0)$ is defined as:

$$KDD \equiv I(\theta_0, \theta_k) = E_{\theta_0}\{\ln \frac{h(Y|\theta_0)}{h(Y|\theta_k)}\}. \quad (1)$$

where E_θ is the expected value under $h(Y|\theta)$. Several model selection criteria are derived from this approach, but the most common are AIC, AIC_c, AIC_{sur}, and $cAIC$. We decribe them below

a. Akaike Information Criterion (AIC). The Akaike Information Criterion (Akaike, 1973), is define as:

$$AIC = -2\log L + 2k, \quad (2)$$

where L is the likelihood and $2k$ represents a penalty that is equivalent to two times the number of model parameters. In case of Cox model, the likelihood is replaced by partial likelihood.

AIC, which stands for Akaike Information Criterion, is considered an asymptotically unbiased estimate of the awaited KDD value (Posada and Buckley, 2004). It was the first widely accepted model selection factor and is used in various modeling frameworks for model selection and evaluation across diverse fields of study (Akpa and Unuabonah (2011), Cavanaugh and Neath (2019)). However, AIC has some limitations. While it is asymptotically effective, it is not strongly consistent (Shibata (1980), Shibata (1981), Claeskens et al. (2008)). AIC can also tend to choose overfitted models more easily (Acion, 2011). Additionally, it is known to be unstable and may fail to consider stochastic errors inherited during the variable selection step (Du et al., 2010). Claeskens and Hjort (2003) demonstrated that while AIC aims to choose a model with good overall properties, the preferred model may not always be good for predicting specific parameters. In the presence of censored observations, the adequacy of AIC may need to be considered (dos Santos Junior and Schneider, 2022). Naik et al. (2007) also reported that AIC can overestimate the number of components for a mixture model, leading to the retention of incorrect variables. Furthermore, in cases of limited sample size, AIC can favor models with excessively high parameters, limiting its effectiveness Hurvich and Tsai (1989).

b. Corrected Akaike Information Criterion (AIC_c). The corrected Akaike Information Criterion (AIC_c) provides an almost unbiased estimation of the Kullback-Leibler information. For biological data in the real world, where parameter estimates and understanding underlying processes are the primary objectives rather than prediction, theoretical evidence supports the use of AIC_c. Furthermore, AIC_c tend to select a better model than AIC in cases of low sample size (Hurvich and Tsai, 1995). Hurvich and Tsai (1995) demonstrated that this should be employed in small samples unless the ratio of total observations to parameter number (n/k) is greater than 40. Regardless of the number of observations, Burnham and Anderson (2004) recommended using AIC_c instead of AIC. They proposed a corrected

form of AIC in which the penalty term is replaced by $\frac{2n(k+1)}{n-(k+2)}$. For the Cox model, Grambsch (1998) proposed replacing n with r (number of occurrence). Unfortunately, the bias of the AIC is not entirely reduced by the AIC_c to $O(n^{-2})$ (Imori et al., 2011). it is worth noting that the use of AIC instead of AIC_c is a common error in research (Burnham and Anderson, 2004). The AIC_c is as follows:

$$AIC_c = -2\log L + \frac{2r(k+1)}{r-(k+2)}, \tag{3}$$

with L being the likelihood function.

c. Improved Akaike Information Criterion (AIC_{sur}). Su and Tsai (2006) demonstrated the superiority of the improved Akaike Information Criterion over the AIC for the AFT model with small sample sizes. Liang and Zou (2008) showed that AIC_{sur} is the best model selection criterion regardless of sample sizes and variances compared to AIC in accelerated failure time models. They also stated that using the number of occurrences would boost the superiority of AIC_{sur} over AIC in situations where the sample size was large. However, this criterion should only be used for parametric models (Liang and Zou, 2008). Liang and Zou (2008) proposed an improved AIC as follows:

$$AIC_{sur} = AIC + \frac{2(k+2)(k+3)}{n-k-3}, \tag{4}$$

with n and k represent the observations and parameter number respectively.

d. Conditional Akaike Information Criterion ($cAIC$). Vaida and Blanchard (2005) claimed that AIC and AIC_c are unacceptable in mixed effects models when the research focuses on clusters. They suggested the $cAIC$. Thereby, this criterion has been enlarged for proportional hazards mixed models. In addition, Ha et al. (2007) suggested a $cAIC$ for frailty models based on hierarchical likelihood; the models they investigated comprised just random intercepts. Moreover, Donohue et al. (2011) adapted the traditional AIC to Parametric mixed models. The conditional AIC as:

$$cAIC = -2\log L + 2\rho, \tag{5}$$

where L is the conditional likelihood and ρ the degree of freedom.

3.2.2. Criteria derived from Kullback's Symmetric Divergences

Kullback's Symmetric Divergence is then defined as:

$$KSD \equiv J(\theta_0, \theta_k) = I(\theta_0, \theta_k) + I(\theta_k, \theta_0). \tag{6}$$

Remark that $J(\theta_0, \theta_k) = J(\theta_k, \theta_0)$, whereas $I(\theta_0, \theta_k) \neq I(\theta_k, \theta_0)$ unless $\theta_k = \theta_0$; thus $J(\theta_0, \theta_k)$ is symmetric. From this approach derives model selection criteria. These

criteria were not discussed in survival analysis, but they were proved in non-linear and generalized linear models (Kim and Cavanaugh, 2005; Kim et al., 2014). We describe them below.

A. Kullback Information Criterion. According to Cavanaugh and Noe (1999), the Kullback Information Criterion (KIC) is as follows:

$$KIC = -2\log L + 3k, \tag{7}$$

k denotes the parameters number, while L denotes likelihood function.

B. Corrected Kullback Information Criterion. Analogously, Cavanaugh (2004) suggests KIC_c):

$$KIC_c = -2\log L + \frac{((k+1)(3n-k-2)}{(n-k-2)}, \tag{8}$$

where k is the parameters number, n is sample size, and L is likelihood. Referring to equation 3, we have:

$$KIC_c = -2\log L + \frac{((k+1)(3r-k-2)}{(r-k-2)}, \tag{9}$$

where r is the number of uncensored events.

3.3. Comparison of criteria

The model selection tools available in survival analysis have been designed from different perspectives. AIC represents a parsimonious method for calculating the predicted K-L distance (Bumham and Anderson, 2002). The AIC provides an efficient way to choose the best approximate model compared to the true model. When the model is too complex to be estimated parametrically, AIC is preferable (Shao, 1997). However, AIC select the overfiting models in case of low sample size compared to parameter number (Rao et al., 2001). AIC has a preference for larger models and cannot effectively select a model among models derived from different data sources (Bumham and Anderson, 2002). Consequently, Sugiura (1978) introduced a corrected version of AIC known as AIC_c. Hurvich and Tsai (1989) demonstrated that corrected Akaike Information Criterion outperformed AIC for small samples. Bumham and Anderson (2002) recommend using $AICc$ when $n/k \leq 40$. When n/k is sufficiently large, AIC and AIC_c are equivalent and likely to select an identical model (Lee and Ghosh, 2009). In the same context, the Improved AIC has been proposed (Liang and Zou, 2008). However, Improved AIC is only used for parametric models (Liang and Zou, 2008). Furthermore, AIC_c values can be easily calculated manually (Bumham and Anderson, 2002).

Another criterion is the Takeuchi Information Criterion (Takeuchi, 1976). The Takeuchi Information Criterion (TIC) outperforms the AIC when the sample size is large enough (Bumham and Anderson, 2002). In geographically weighted Cox regression models, the modified TIC is employed rather than the AIC to be accounted for the trade-off between variance and bias in addition to partial likelihood maximization (Xue et al., 2020; Seidi et al., 2023). However, Shibata (1989) notes that the estimation error of the two first partial derivatives of the likelihood of the TIC can lead to instability in the model selection results. For this reason, Bumham and Anderson (2002) advises against using it unless the sample size tends to infinity and good estimations of the partial two-first derivatives likelihood may be expected. According to Acion (2011), the TIC is not as well-known among practitioners compared to the AIC. This is primarily due to the fact that the TIC was published in a difficult-to-find Japanese newspaper.

Schwarz (1978) is credited with developing the Bayesian Information Criterion (BIC). However, it is important to note that the BIC is not an estimate of expected relative K-L information. Generally, the models selected by Bayesian Information Criterion tend to be more parsimonious compared to those selected by the Akaike Information Criterion (Lee and Ghosh, 2009). The BIC has a higher likelihood of selecting the true model compared to them AIC (Kuk and Varadhan, 2013). It is particularly effective when the exact model is one of the potential models under consideration (Vrieze, 2012). However, it is worth mentioning that the BIC has a tendency to favor smaller models (Kuk and Varadhan, 2013). In scenarios with finite sample sizes, the BIC may be less efficient than the AIC, even when the true model is taken into account (Vrieze, 2012).

Claeskens and Hjort (2003) proposed a Focused Information Criterion (FIC). It is a frequentist approach which select various models to distinct objectives. FIC reduces the mean square error of the specific parameter, while AIC or BIC, try to select a more parsimonious model (Behl et al., 2012). Thus, the FIC does not attempt to evaluate the goodness of fit of the potential models, but instead concentrates on the major parameter of concern (Yang et al., 2015).

The Kullback Information Criterion was suggested deriving from KSD Cavanaugh and Noe (1999). KIC_c is a corrected version of KIC_c Cavanaugh (2004).

3.4. On the usage of criteria in survival analysis

Several model selection criteria were considered in the literature, including the AIC, the AIC_c, the AIC_{sur}, and the $cAIC$ (Figure 5(b), page 3390). The most used criteria were the AIC(94.62 %) and the AIC_c (3.48 %). The most widespread use of the AIC can be attributed to the fact that it is the first widely accepted criterion for model selection (Cavanaugh and Neath, 2019). Another reason is that AIC is easy and quick to calculate (Auranen et al., 2005), compares numerous nested or non-nested models simultaneously, and evaluates model selection uncertainty (Posada and Buckley, 2004). None of the papers employs the criteria derived from

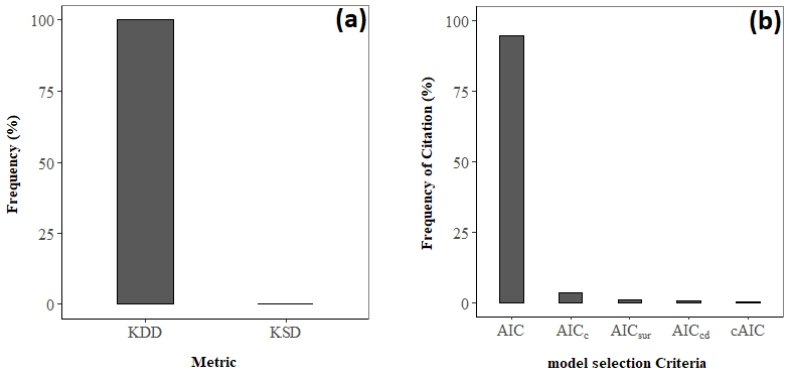

Figure 5. Distribution of studied papers according to a (a) metric that serves as the basis for the model selection criteria, (b) model selection criteria.

KSD, such as KIC, KIC_c, and KIC_{sur} (Figure 5(a), page 3390). However, Kullback's Symmetric Divergence based selection criteria were found to outperform those derived from Kullback's Directed Divergence (AIC, AIC_c, AIC_{sur}, $cAIC$, etc.) for linear regression (Cavanaugh, 2004), longitudinal data (Azzaoui and Hafidi, 2012), and overdispersed count data (Kim et al., 2014).

Several types of models were employed in the selected papers to analyze the influence of covariates on survival time. Some were extremely simple, while others were more complex. The Cox proportional hazards (PH) model (48.95 %), Parametric PH model (22.51 %), Accelerated Failure Time (AFT) model (9.95 %), and frailty models (8.11 %) were the most commonly used in survival analysis (Figure 6,page 3391). The Cox PH model was preferred since it made no assumptions about survival times distribution (Kumar et al., 2020). The parametric AFT model serves as an alternative to the PH model, not relying on the proportional hazard assumption (Aida et al., 2022). The frailty models employed in the papers included Proportional Hazards frailty (70.83 %), Accelerated Failure Time frailty (16.67 %), and Variance Corrected frailty Proportional Hazard models (12.5 %). Furthermore, the random variable in frailty models was associated with distributions such as Gamma (48.98 %), Inverse Gaussian (36.73 %), Positive Stable (6.12 %), Compound Poisson (4.08 %), Gaussian (2.04 %), and Generalized Gamma (2.04 %).

A total of 59 different distribution types were identified in the chosen papers. Among these, the most common distributions used were the Weibull (21.81 %), log-Normal (16.87 %), log-logistic (16.46 %), and Exponential (14.81 %) (Figure 7, page 3391). The distributions used also depended significantly on the model, as evidenced by the p-value of Fisher's exact test ($p = 0.0005$).

Right censoring (95.39 %) and interval censoring were the most frequently utilized in the selected papers (Table 1, 3392). It was revealed that papers using frailty

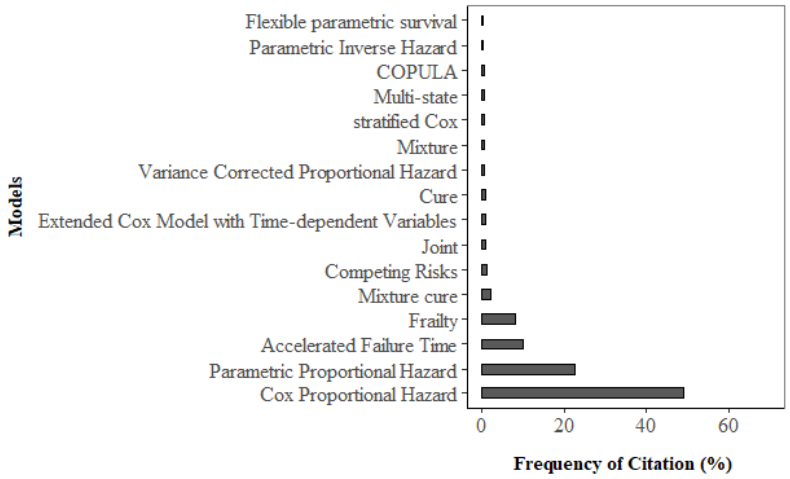

Figure 6. Frequency of models used in survival analysis.

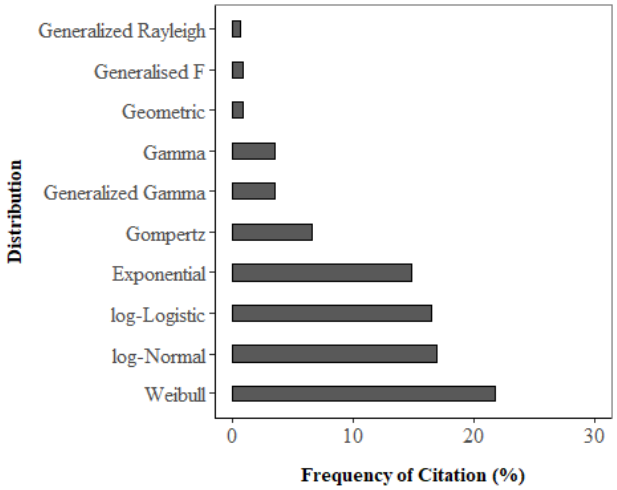

Figure 7. Frequency of the ten most used distributions.

models did not use $cAIC$ or $mAIC$ to select models. The ratio of observations to parameters, on the other hand, was less than 40 in 35.12 % of the papers that used AIC (Figure 8, page 3392). Furthermore, this ratio was greater than 40 in 20 % and 33.33 % of those who used AIC_c and AIC_{sur}, respectively. We found that AIC was significantly more frequently utilized, regardless of the model, even though (Xu et al., 2009) and (Donohue et al., 2011) advocate using the marginal AIC and conditional AIC, respectively, for the mixed proportional hazard model, including the frailty model. Additionally, it is advised to use the AIC_c or AIC_{sur} when the number of parameters is large in comparison with the number of observations or

Table 1. Number of papers according to the types of censoring

Type of censoring and truncation	Number of papers (percentage)
Interval censoring	7 (2.30%)
Left censoring	3 (0.99%)
Right censoring	290 (95.39%)
Truncation	1 (0.33%)
Not reported	3 (0.99%)

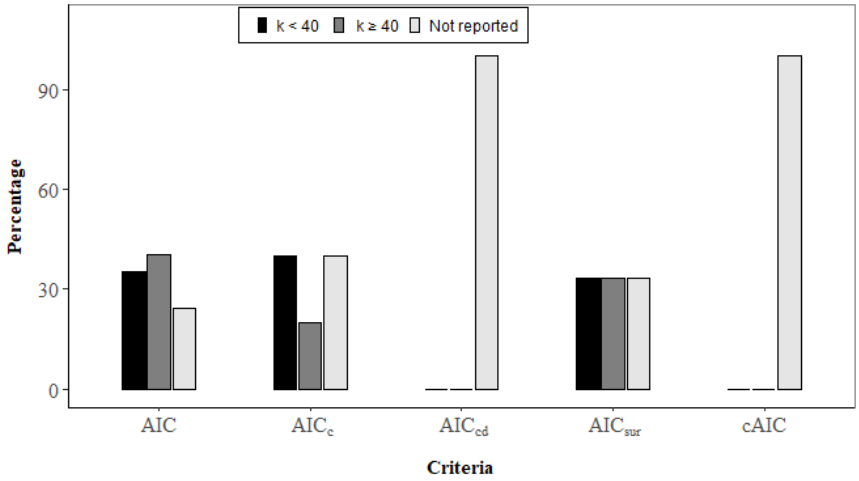

Figure 8. The frequency of model selection criteria is according to the ratio of observations and parameters. $K = n/p$ with n = number of observations and p = number of parameters.

when or when $n/p < 40$ (Hurvich and Tsai, 1995). Therefore, the AIC_c or AIC_{sur} is used for low sample size. It is important to note that over one-third of papers using AIC should use AIC_c or AIC_{sur}. Furthermore, the AIC should be used in about half of the studies that chose models using AIC_c and AIC_{sur}.

4. Gaps and perspectives

The majority of the included studies used AIC for model selection. However, many of these studies should have considered the model they were using when choosing a model selection criterion. This is especially true for studies that used the frailty model (e.g. Hanagal and Dabade, 2014; Banbeta et al., 2015; Gurmu, 2018; Sidhu et al., 2019). Vaida and Blanchard (2005) demonstrated that AIC is unacceptable in mixed effects models when the investigation focuses on clusters. Additionally, the conventional model selection criterion AIC in a mixed effect model uses the marginal likelihood (Liang et al., 2008). However, many of the selected publications that use the frailty model calculate AIC using the likelihood rather than the

marginal likelihood (e.g. Ghadimi et al., 2012; Chilot et al., 2022). The same observation of the use of likelihood is made in papers that implemented the Cox model, e.g. Pourhoseingholi et al., 2007; Ebrahimi et al., 2019; Chilot et al., 2022), though Cox (1975) recommends partial likelihood in a Cox model. But, statistical software calculates the AIC using marginal and partial likelihood for the frailty and Cox models, respectively. Furthermore, while choosing the criterion, the sample size or the ratio n/p is not considered. Improved AIC is recommended for models of an accelerated lifetime that use Exponential, Weibull, Log-Logistic, Log-Normal, and Generalized Gamma distributions, as advised by Liang and Zou (2008). Furthermore, for AIC_c or AIC_{sur} in the penalty term, it is recommended to use the events number rather than the observations number (Grambsch, 1998; Liang and Zou, 2008). However, these recommendations are rarely followed (e.g. Park and Qiu, 2014; Barbu et al., 2020), Montaseri et al. (2016)). Finally, no paper discusses the new criteria (derived from KSD), which were claimed to be superior in linear, non-linear, and generalized linear models (Cavanaugh, 2004; Kim and Cavanaugh, 2005; Azzaoui and Hafidi, 2012; Kim et al., 2014). Future studies could assess the effectiveness of criteria derived from KSD. As a result, the performance of criteria for the most commonly used survival models, including the Cox PH, the AFT, and the Frailty models, could be evaluated. In this way, researchers could better understand that choosing the proper criterion for selecting the true model varies depending on the context. In addition, interactions between variables are much more common in public health, epidemiology, and clinical trials (Rod et al., 2012). A future study could look into how covariate interactions affect the robustness of model selection criteria. Implementing different model selection criteria derived from KSD in survival analysis software packages (such as R or SAS) is also of great importance for future research.

5. Conclusion

This study presents a systematic review of model selection criteria in survival analysis, focusing on those based on Kullback's divergence. The analysis revealed several key findings. Survival analysis, with its significant importance in recent years, utilizes various model selection criteria such as AIC, AIC_c, AIC_{sur}, and $cAIC$. However, we observed that none of the criteria derived from Kullback's Symmetric Divergence are commonly employed in this field. Furthermore, it was noted that researchers often lack consistent adherence to minimal guidelines when selecting a model selection criterion. Based on the findings of this study, it is crucial for researchers to carefully consider the appropriate criteria for selecting the most accurate model. Additionally, evaluating the performance of criteria derived from Kullback's Symmetric Divergence in survival models is essential for future research.

Acknowledgments. We are grateful for the financial assistance provided by the German Academic Exchange Service and the African Excellence Center in Mathematical Sciences and Applications of the Institute of Mathematics and

Physical Sciences in Benin, as well as their respective backers.

We would also like to thank the members of the Editorial Board and the anonymous referees for taking the necessary time and effort to review the manuscript. We sincerely appreciate all your valuable comments and suggestions, which helped us in improving the quality of the manuscript.

References

Acion. C. L., 2011. *Criteria for generalized linear model selection based on Kullback's symmetric divergence*. PhD thesis, University of Iowa.

Aida. H, Hayashi. K, Takeuchi. A, Sugiyama. D, and Okamura. T. , 2022. An accelerated failure time cure model with shifted gamma frailty and its application to epidemiological research. In *Healthcare*, volume 10, page 1383. MDPI.

Akaike. H., 1973. Information theory and an extension of the maximum likelihood principle. In *Proc. 2nd International Symposium on Information Theory, 1973*, pages 267–281. Akademiai Kiado.

Akaike. H. , 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Akpa. O. M. and Unuabonah. E. I., 2011. Small-sample corrected akaike information criterion: an appropriate statistical tool for ranking of adsorption isotherm models. *Desalination*, 272(1-3):20–26.

Auranen. T., Nummenmaa. A., Hämäläinen. M. S., Jääskeläinen. I. P., Lampinen. J., Vehtari. A., and Sams.M., 2005. Bayesian analysis of the neuromagnetic inverse problem with $\ell p - norm$ priors. *NeuroImage*, 26(3):870–884.

Azzaoui. N. and Hafidi. B., 2012. Criteria for longitudinal data model selection based on kullback's symmetric divergence. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, 15.

Banbeta. A., Seyoum. D., Belachew. T., Birlie. B., and Getachew. Y., 2015. Modeling time-to-cure from severe acute malnutrition: application of various parametric frailty models. *Archives of Public Health*, 73(1):1–8.

Barbu. V. S. , Karagrigoriou. A., and Makrides. A. . , 2020. Statistical inference for a general class of distributions with time-varying parameters. *Journal of Applied Statistics*, 47(13-15):2354–2373.

Behl. P., Dette. H., Frondel. M., and Tauchmann. H., 2012. Choice is suffering: A focused information criterion for model selection. *Economic Modelling*, 29(3): 817–822.

Bumham. K. P. and Anderson. D. R., 2002. Model selection and multimodel inference: a practical information-theoretic approach. *Spnnger-Veflag, New York, New York*.

Bumham. K. P. and Anderson. D. R., 2004. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Cavanaugh. J. E., 2004. Criteria for linear model selection based on kullback's symmetric divergence. *Australian & New Zealand Journal of Statistics*, 46(2): 257–274.

Cavanaugh. J. E. and Neath. A. A., 2019. The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3):e1460.

Cavanaugh. M. A. and Noe. R. A., 1999. Antecedents and consequences of relational components of the new psychological contract. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 20(3):323–340.

Chilot. D., Belay. D. G., Shitu. K., Mulat. B., Alem.A. Z., and Geberu. D. M., 2022. Prevalence and associated factors of common childhood illnesses in sub-saharan africa from 2010 to 2020: a cross-sectional study. *BMJ open*, 12(11):e065257.

Claeskens. G. and Hjort. N. L., 2003. The focused information criterion. *Journal of the American Statistical Association*, 98(464):900–916.

Claeskens. G. and Hjort. N. L., et al., 2008. Model selection and model averaging. *Cambridge Books*.

Cox. D., 1972. Regression models and life tables," journal of the royal statistical society, series b, 34, 187-220..(1975). *Partial Likelihood," Biometrika*, pages 62–269.

Cox. D. R., 1975. Partial likelihood. *Biometrika*, 62(2):269–276.

Donohue. M., Overholser. R., Xu. R., and Vaida. F., 2011. Conditional akaike information under generalized linear and proportional hazards mixed models. *Biometrika*, 98(3):685–700.

dos Santos Junior. P. C. and Schneider. S., 2022. Power piecewise exponential model for interval-censored data. *Journal of Statistical Theory and Practice*, 16 (2):26.

Du. P., Ma.S., and Liang. H., 2010. Penalized variable selection procedure for cox models with semiparametric relative risk. *The Annals of Statistics*.

Ebrahimi. V., Khademian. M. H., Masoumi. S. J., Morvaridi. M. R. , and Ezzatzadegan Jahromi. S., 2019. Factors influencing survival time of hemodialysis patients; time to event analysis using parametric models: a cohort study. *BMC nephrology*, 20(1):1–9.

Ghadimi. M. R., Mahmoodi. M., Mohammad. K., Rasouli. M., Zeraati. H., and A. Fotouhi., 2012. Factors affecting survival of patients with oesophageal cancer: a study using inverse gaussian frailty models. *Singapore medical journal*, 53(5): 336.

Grambsch. T. M. T. P. M., 1998. Penalized cox models and frailty. *echnical report, Division of Biostatistics.*, (12):156–175.

Gurmu. S. E., 2018. Assessing survival time of women with cervical cancer using various parametric frailty models: a case study at tikur anbessa specialized hospital, addis ababa, ethiopia. *Annals of Data Science*, 5(4):513–527.

Ha. I. D., Lee. Y., and MacKenzie. G., 2007. Model selection for multi-component frailty models. *Statistics in Medicine*, 26(26):4790–4807.

Hanagal. D. D. and Dabade. A. D., 2014. Comparisons of frailty models for kidney infection data under weibull baseline distribution. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(4):342–373.

Hurvich. C. M. and Tsai. C.-L., 1989. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Hurvich. C. M. and Tsai. C.-L., 1995. Model selection for extended quasi-likelihood models in small samples. *Biometrics*, pages 1077–1084.

S. Imori, H. Yanagihara, and H. Wakaki. General formula of bias-corrected aic in generalized linear models. Technical report, TR, 2011.

H.-J. Kim and J. E. Cavanaugh. Model selection criteria based on kullback information measures for nonlinear regression. *Journal of statistical planning and inference*, 134(2):332–349, 2005.

H.-J. Kim, J. E. Cavanaugh, T. A. Dallas, and S. A. Foré. Model selection criteria for overdispersed data and their application to the characterization of a host-parasite relationship. *Environmental and ecological statistics*, 21(2):329–350, 2014.

D. Kuk and R. Varadhan. Model selection in competing risks regression. *Statistics in medicine*, 32(18):3077–3088, 2013.

S. Kullback. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243, 1968.

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

Lawless. J. F., 2011. *Statistical models and methods for lifetime data*. John Wiley & Sons.

Lee. H. and Ghosh. S. K., 2009. Performance of information criteria for spatial models. *Journal of statistical computation and simulation*, 79(1):93–106.

Liang. H. and Zou. G., 2008. Improved aic selection strategy for survival analysis. *Computational statistics & data analysis*, 52(5):2538–2548.

Liang. H., Wu. H., and Zou. G., 2008. A note on conditional aic for linear mixed-effects models. *Biometrika*, 95(3):773–778.

Lu. C.-L., Wang.S., Ji. Z., Wu. Y., Xiong. L., Jiang.X., and Ohno-Machado.L., 2015. Webdisco: a web service for distributed cox model learning without patient-level data sharing. *Journal of the American Medical Informatics Association*, 22(6): 1212–1219.

Montaseri. M., Charati. J. Y., and Espahbodi. F., 2016. Application of parametric models to a survival analysis of hemodialysis patients. *Nephro-urology monthly*, 8(6).

Naik. P. A., Shi. P., and Tsai. C.-L., 2007. Extending the akaike information criterion to mixture regression models. *Journal of the American Statistical Association*, 102(477):244–254.

Page M. J., McKenzie. J. E., Bossuyt. P. M., Boutron. I., Hoffmann. T. C., Mulrow. C. D., Shamseer. L., Tetzlaff. J. M., Akl. E. A., Brennan. S. E., et al., 2021. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews*, 10(1):1–11.

Park. K. Y. and Qiu. P., 2014. Model selection and diagnostics for joint modeling of survival and longitudinal data with crossing hazard rate functions. *Statistics in medicine*, 33(26):4532–4546.

Posada. D. and Buckley. T. R., 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic biology*, 53(5):793–808.

Pourhoseingholi. M. A., Hajizadeh. E., Moghimi Dehkordi. B., Safaee. A., Abadi.A., Zali. M. R., et al., 2007. Comparing cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pacific Journal of Cancer Prevention*, 8(3):412.

Ramasamy. R. and Kaliannan. M., 2021. Monitoring the newly infected cases of covid-19 data weekly: A survival data analysis (sda) perspective. *Statistical Journal of the IAOS*, (Preprint):1–16.

Rao. C., Wu. Y., Konishi.S., and Mukerjee. R., 2001. On model selection. *Lecture Notes-Monograph Series*, pages 1–64.

RodN. H., Lange. T., Andersen. I., Marott. J. L., and Diderichsen. F., 2012. Additive interaction in survival analysis: use of the additive hazards model. *Epidemiology*, 23(5):733–737.

Schober. P. and Vetter. T. R., 2018. Survival analysis and interpretation of time-to-event data: the tortoise and the hare. *Anesthesia and analgesia*, 127(3):792.

G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

Seidi . N., Tripathy. A., and Das. S. K., 2023. Using geographic location-based public health features in survival analysis. *arXiv preprint arXiv:2304.07679*.

Shaik. A. B., Venkataramanaiah. M., and Thasleema. S., 2015. Statistical applications of survival data analysis for breast cancer data. *i-Manager's Journal on Mathematics*, 4(2):30.

Shao. J., 1997. An asymptotic theory for linear model selection. *Statistica sinica*, pages 221–242.

Shibata. R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The annals of statistics*, pages 147–164.

Shibata. R., 1981. An optimal selection of regression variables. *Biometrika*, 68(1): 45–54.

Shibata. R., 1989. *Statistical aspects of model selection*. Springer.

Sidhu. S., Jain. K., and Sharma. S. K., 2019. The generalized gamma shared frailty model under different baseline distributions. *International Journal of Mathematical, Engineering and Management Sciences*, 4(1):219.

Su. X. and Tsai. C.-L., 2006. An improved akaike information criterion for generalized log-gamma regression models. *The International Journal of Biostatistics*, 2 (1).

Sugiura. N., 1978. Further analysis of the data by akaike's information criterion and the finite corrections: Further analysis of the data by akaike's. *Communications in Statistics-theory and Methods*, 7(1):13–26.

Takeuchi. K., 1976. Distribution of an information statistic and the criterion for the optimal model. *Mathematical Science*, 153:12–18.

Vaida. F. and Blanchard. S., 2005. Conditional akaike information for mixed effects models. *Corrado Lagazio, Marco Marchi (Eds)*, page 101.

Vrieze. S. I., 2012. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228.

Wei. L.-J. , 1992. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine*, 11(14-15):1871–1879.

Xu. R., Vaida. F., and Harrington. D. P., 2009. Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models. *Statistica Sinica*, 19(2):819.

Xue. Y., Schifano. E. D., and Hu. G., 2020. Geographically weighted cox regression for prostate cancer survival data in louisiana. *Geographical Analysis*, 52(4):570–587.

Yang. H., Liu. Y., and Liang. H., 2015. Focused information criterion on predictive models in personalized medicine. *Biometrical Journal*, 57(3):422–440.