

Evaluating the Impact of Baseline Hazard Function Misspecification on Treatment Effect Estimation

Charly Fowler, Hanfei Qi, Robert Tumasian III , and Haoyang Yi

February 24, 2021

Objectives

The goal of this study is to evaluate how misspecifying the baseline hazard function can influence the estimation of treatment effects in survival without censored observations (complete data available for all subjects; no missingness due to dropout or event occurrence). This work focuses on conducting simulations to compare the exponential and Weibull proportional hazards models to the Cox proportional hazards model. We also discuss the impact of utilizing an overly complicated model (e.g., Cox) when a less complex model (e.g., exponential) is sufficient.

Statistical Methods

Survival analysis is used to analyze time-to-event data (e.g., time to symptom onset or time to mortality). Survival functions, $S(t)$, measure the probability of an individual not experiencing an event beyond a certain time t . Similarly, hazard functions, $h(t)$, measure the instantaneous risk of failure at a certain time t , given that the individual has not experienced an event until that time. The hazard function can be expressed as $\frac{f(t)}{S(t)}$, where $f(t)$ is the distribution of survival times.

One purpose of proportional hazards modeling is to assess the effectiveness of a particular treatment (X) over survival time T , where the hazard ratio for patient i at time t is defined as $h_i(t) = h_0(t)e^{x_i\beta}$. Here, $h_0(t)$ denotes the pre-specified baseline hazard function, x_i indicates treatment allocation (0=control, 1=treatment), and β represents the log hazard ratio, or the hazard reduction among treated individuals compared to the control group. Thus, the proportional hazard can be expressed as $\frac{h(t|x_0)}{h(t|x_1)} = e^{\beta(x_0-x_1)}$, which is independent of survival time t .

Simulation Design

We consider three proportional hazards models (exponential, Weibull, and Cox) and six baseline hazard functions (exponential [$\gamma = 1$], Weibull [$\gamma = 0.5, 2, 5$], lognormal, and piecewise). Note that the exponential and Weibull models are equivalent when $\gamma = 1$. All continuous baseline hazard functions are provided in Table 1 below.

	Exponential	Weibull	Lognormal
$h_0(t)$	λ	$\lambda\gamma t^{\gamma-1}$	$\frac{1}{t\sigma} \phi\left(\frac{\ln(t)}{\sigma}\right)$ $\Phi\left(\frac{-\ln(t)}{\sigma}\right)$

Table 1. Continuous baseline hazard functions, $h_0(t)$, considered in this study. The normal PDF and CDF are denoted by ϕ and Φ , respectively.

The exponential and Weibull models can implement different baseline hazard functions, while the Cox model estimates β without this specification. It is important to mention that even if the baseline hazard function is known, the Cox model is still expected to perform well due to its semi-parametric efficiency (Anderson et al, 1982). All three models impose the restraint that the effect of the treatment be multiplicative on the hazard curve. The baseline hazard curves we consider are shown in Figure 1 below.

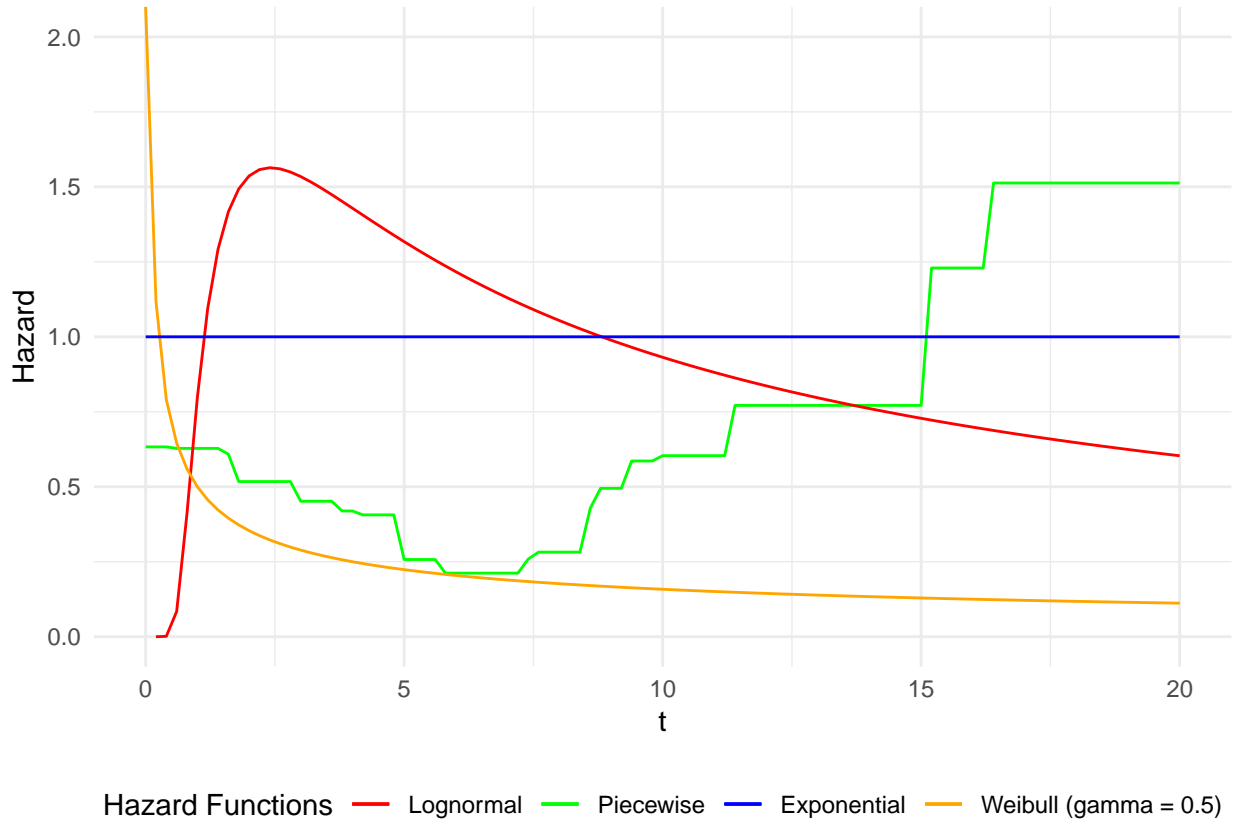


Figure 1: Baseline hazard functions used in this study for data generation. The lognormal distribution uses $\mu = 0$ and $\sigma = 0.5$. The piecewise distribution (CHARLY: ADD HOW WAS THIS MADE). The Weibull distributions with $\gamma = 2, 5$ have been omitted due to scaling (very steep growth).

All simulation data was generated using the `simsurv` function in the `simsurv` package. We defined a binomial treatment variable (`trt`) with $p = 0.5$ to ensure equal likelihood of random assignment to the treatment or control group. All parameters are set according to different distributions of hazards. The resulting dataset contains time of event (`eventtime`), status (`status`) and treatment (`trt`). We did not simulate any censoring, thus the event (`eventtime`) variable represents time of event for all rows.

We simulated survival data from six different baseline hazard distributions. First, we simulated from an exponential distribution with $\lambda = 1$, and from a Weibull with $\lambda = 0.1$, and γ values of 0.5, 2, and 5.

These values of γ were chosen to consider constant, monotone decreasing, linearly increasing, and monotone increasing hazard curves. From these simulations we aim to test the impact of utilizing an overly complex model, such as Cox, when an exponential or Weibull model would suffice.

We additionally simulated lognormal data with $\mu = 0$, $\sigma = 0.5$, which generates a non-monotone hazard curve, to understand the implications of misspecifying the underlying distribution when fitting a model. These first five hazard functions we simulated from can be seen in Table 1, below.

Lastly, we simulated from a piecewise step function hazard curve, inspired by Brilleman et al. This hazard curve is visualized in Figure 1, below. It was included to further explore the implications of fitting an overly-simplistic model such as exponential or Weibull on data with a complex underlying distribution.

When simulating the data, we used $\beta = -0.5$ as the true effect size for all underlying distributions.

We set the sample size as 100 and for each underlying distribution sampled 1000 times to get the dataframe of results. Each time we fitted exponential, Weibull and cox model to the data and extracted the β estimates provided by three models.

Results

To assess model performance, we plotted the survival probability versus time of all models with a sample from each simulation scheme. We also used mean-squared error (MSE) to check model performance.

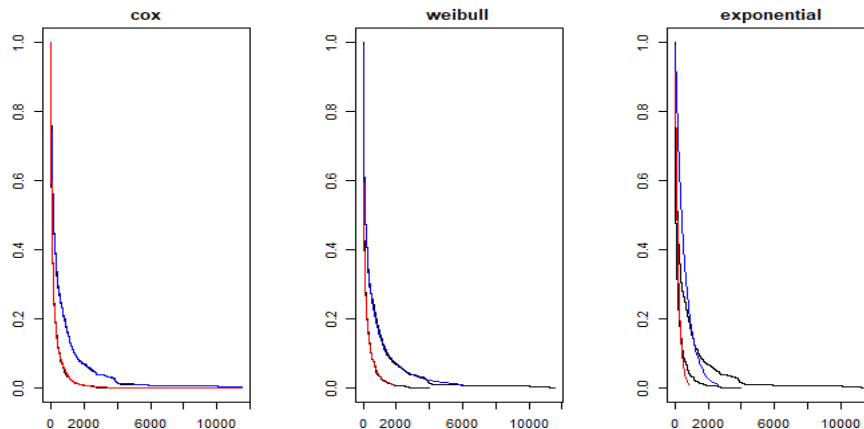


Figure 2. Survival probability versus time of each model, baseline hazard function follows weibull distribution with $\gamma = 0.5$.

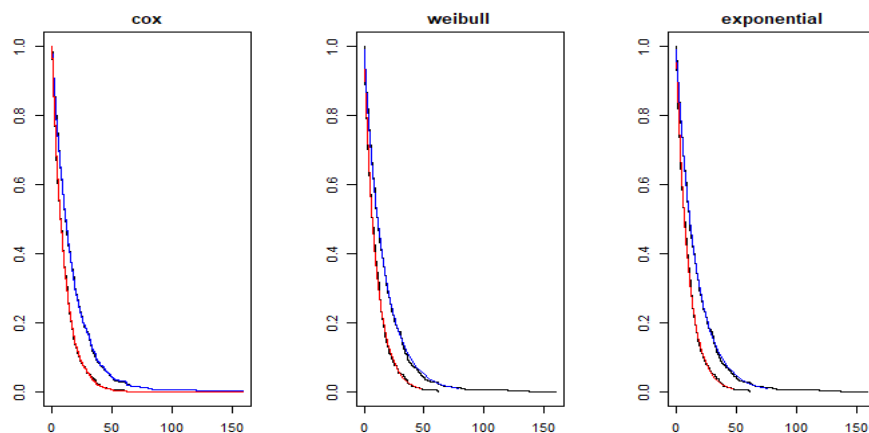


Figure 3. Survival probability versus time of each model, baseline hazard function follows Exponential distribution.

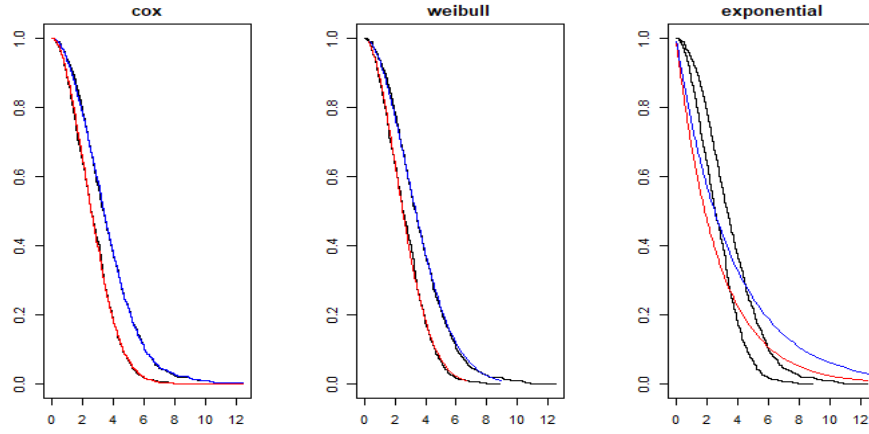


Figure 4. Survival probability versus time of each model, baseline hazard function follows weibull distribution with $\gamma = 2$.

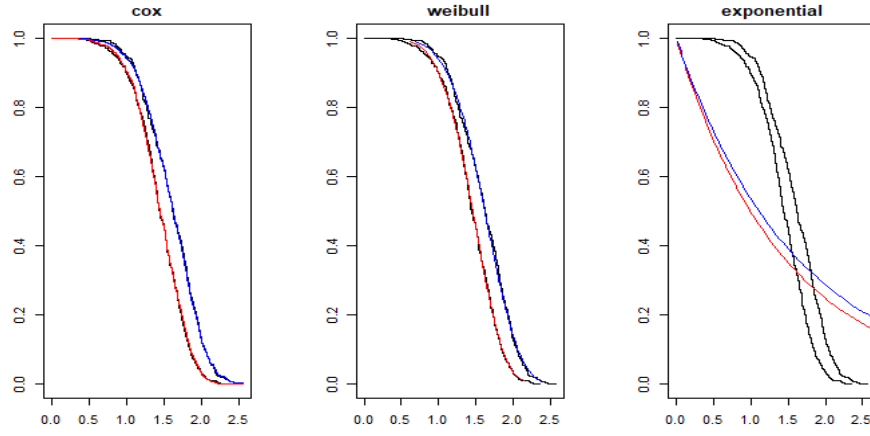


Figure 5. Survival probability versus time of each model, baseline hazard function follows weibull distribution with $\gamma = 5$.

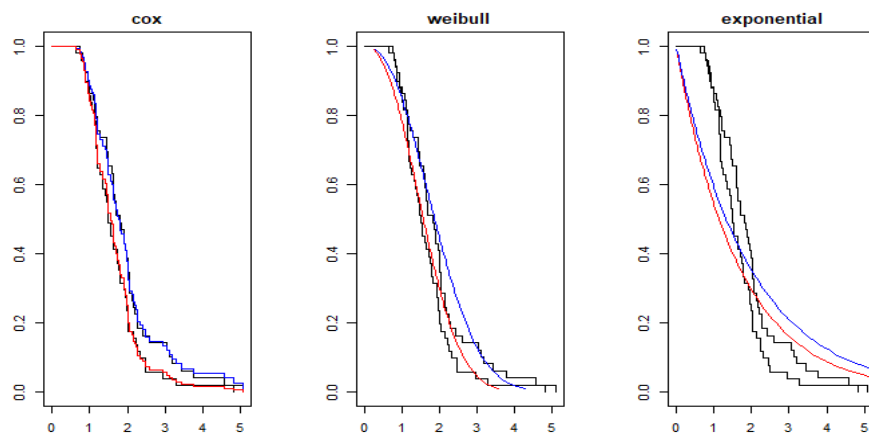


Figure 6. Survival probability versus time of each model, baseline hazard function follows lognormal distribution.

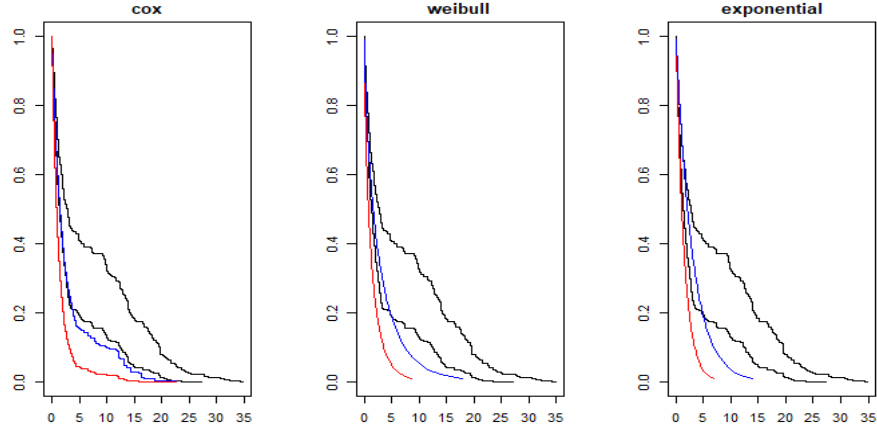


Figure 7. Survival probability versus time of each model, baseline hazard function follows piecewise distribution.

Model	Gamma_0.5	Gamma_1	Gamma_2	Gamma_5	Lognormal	Piecewise
Exponential	0.3297817	0.1499803	0.0463993	0.1615276	0.0728173	0.0558371
Weibull	0.3135206	0.1511192	0.0487907	0.0436235	0.0936464	0.0464780
Cox	0.3091898	0.1489322	0.0489004	0.0461492	0.0422765	0.0422971

Table 2. The MSE table of three models with each simulation data.

Model	Gamma_0.5	Gamma_1	Gamma_2	Gamma_5	Lognormal	Piecewise
Exponential	-0.5473097	-0.5227732	-0.3500854	-0.1006702	-0.2481725	-0.5759623
Weibull	-0.5328547	-0.5240389	-0.5044208	-0.5087976	-0.6558937	-0.5281895
Cox	-0.5290293	-0.5201727	-0.5010552	-0.5065809	-0.5104640	-0.5104477

Table 3. The Mean table of three models with each simulation data.

Conclusions

By comparing the Mean square error(MSE) and mean of β derived from three models in different simulations, we found that misspecifying the baseline hazard function can lead to significant differences between actual

value of β and the truth(which was set as -0.5).

For example, we got high MSEs and biased means of β when applying exponential and Weibull model in data simulated by lognormal distributions, which indicates that the treatment effect was not evaluated well. We further confirmed the deviated estimation through plots of survival probability versus time where Weibull and exponential model can not provide the prediction curve that match the actual. In this case, Cox model provided small MSEs, while means of β are quite close to the truth. Also the prediction curve of Cox model was quite consistent with the actual. Thus we concluded that Cox model can estimate the treatment effect with high accuracy and precision which avoids the impacts of misspecifying the baseline hazard function.

Case of the impact of fitting too complicated a model when an exponential is sufficient was found in simulation that $\gamma = 1$. In this case all three models had good performance in MSE and mean of β with slight difference, which indicates that exponential model is sufficient. By observing the image, we noticed that the cox model had over-fitting to the data with low survival probability(the curve extended to $T = 150$), which may lead to a greater impact of end time data (such as censored data) in practical applications. Exponential prediction curve only extended to $T = 50$, which indicates that exponential model can avoid this impact since it has weaker dependence on end time data.

Contributions

Charly Fowler worked on simulation functions, performed piecewise simulation, plotted baseline hazard curves, and created datasets of results. Hanfei Qi worked on formatting simulation functions and editing plotting functions. Robert Tumasian III worked on editing simulation functions, performing lognormal simulation, and creating datasets of results. Haoyang Yi worked on creating plotting functions and editing simulation functions. All members contributed equally to this project.

References

- Andersen, Per Kragh, and Richard D. Gill. "Cox's regression model for counting processes: a large sample study." *The annals of statistics* (1982): 1100-1120.
- Brilleman, Samuel, Rory Wolfe, Margarita Moreno-Betancur, & Michael J. Crowther. "Simulating Survival Data Using the simsurv R Package." *Journal of Statistical Software* [Online], 97.3 (2021): 1 - 27. Web. 23 Feb. 2021