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. 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.

We consider three proportional hazards models (exponential, Weibull, and Cox). 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.

Simulation Design

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. The resulting dataset contains time of event (`eventtime`), status (`status`), and treatment group (`trt`). Since we did not simulate censored observations, the `eventtime` variable represents time of event for all subjects (complete data available for all subjects; no missingness due to dropout or event occurrence).

We simulated survival data from six different baseline hazard distributions, using $\beta = -0.5$ as the true treatment effect size. First, we simulated from an exponential distribution with $\lambda = 0.1$, and then from a Weibull distribution with $\lambda = 0.1$ and $\gamma = 0.5, 2, 5$. Note that when $\gamma = 1$, the exponential and Weibull distributions are equivalent. These baseline hazards were chosen to consider monotone decreasing ($\gamma = 0.5$), constant (exponential), and monotone increasing ($\gamma = 2, 5$) baseline hazard curves. Next, we simulated from lognormal ($\mu = 0, \sigma = 0.5$) and piecewise distributions (inspired by Brilleman et al.) to consider non-monotone baseline hazard functions and to better understand the implications of misspecifying the underlying distribution of survival times when fitting a model. All continuous baseline hazard functions are provided in Table 1 below, and baseline hazard functions are plotted beneath in Figure 1.

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

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.

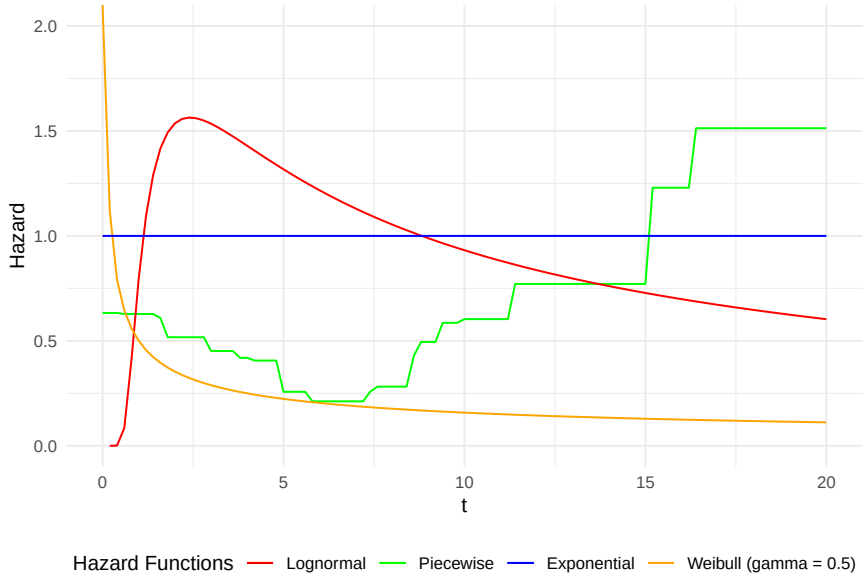

Figure 1. Baseline hazard functions used in this study for data generation. The lognormal distribution uses $\mu = 0$ and $\sigma = 0.5$. The piecewise distribution is based on curve generated by Brilleman et al. The Weibull distributions with $\gamma = 2, 5$ have been omitted due to scaling (very steep growth).

Altogether, 1000 survival datasets containing 100 samples were simulated for each baseline hazard function. Each dataset was used to fit all three proportional hazards models and extract the estimated treatment effects (β). Furthermore, we generated six additional datasets of 100 samples for each baseline hazard function and fit exponential, Weibull, and Cox models for visualization in Figure 2.

Results

To assess model performance, we plotted the survival probability versus time for all three models with a sample from each of the six simulation schemes (Figure 2). We also used mean-squared error (MSE), mean, and variance (?) to check model performance.

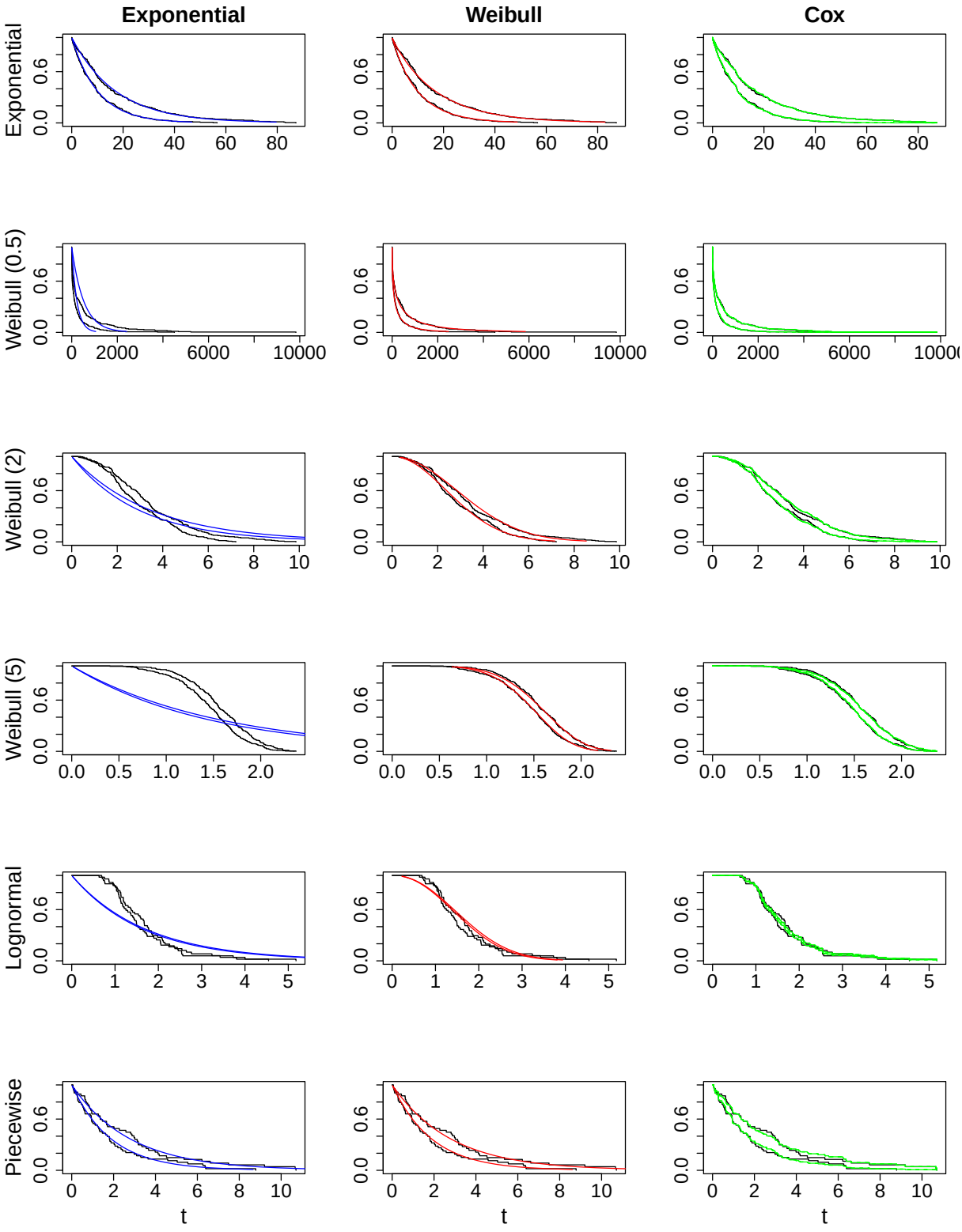

Figure 2. Survival probabilities versus time for all models and baseline hazard functions.

In Figure 2 above, we can observe the impact of baseline hazard misspecification when fitting a proportional hazard model. For example, the exponential model cannot accurately fit a baseline hazard distribution with $\gamma = 5$. For the data with lognormal or piecewise baseline hazard curves, neither the Weibull nor the exponential models follow the datas' survival curve. We also note that the Cox proportional hazard model has good fit across all 6 datasets, even when a simpler model would suffice.

In Table 2, for the six baseline hazard curves we calculate the MSE of the three models estimates for β. The Weibull model had the lowest MSE for data simulated from a Weibull hazard curve with $\gamma = 2, 5$. The Cox proportional hazard model achieves lowest MSE among data simulated from exponential, Weibull with $\gamma = 0.5$, lognormal, and piecewise hazard curves. While the exponential model by MSE never performed best, its performance was nearly identical to the Cox model for data simulated from an exponential distribution. Similarly, for data from a Weibull with $\gamma = 0.5$, the Cox performed only slightly better.

Model	Exponential	Weib_0.5	Weib_2	Weib_5	Lognormal	Piecewise
Exponential	0.1499803	0.3297817	0.0463993	0.1615276	0.0728173	0.0558371
Weibull	0.1511192	0.3135206	0.0487907	0.0436235	0.0936464	0.0464780
Cox	0.1489322	0.3091898	0.0489004	0.0461492	0.0422765	0.0422971

Table 2. The MSE table of three models with each simulation data. Weibull models are identified as `Weib_`γ.

In Table 3, we see the breakdown of the MSE for each simulation and model into mean and variance. (. . .)

Model	Exponential	Weib_0.5	Weib_2	Weib_5	Lognormal	Piecewise
Exponential	-0.5227732	-0.5473097	-0.3500854	-0.1006702	-0.2481725	-0.5759623
Weibull	-0.5240389	-0.5328547	-0.5044208	-0.5087976	-0.6558937	-0.5281895
Cox	-0.5201727	-0.5290293	-0.5010552	-0.5065809	-0.5104640	-0.5104477

Table 3. The Mean (and variance?) table of three models with each simulation data. Weibull models are identified as `Weib_`γ.

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 ($\beta = -0.5$).

For example, the exponential and Weibull models obtained high MSEs and biased means (. . .) of β in data simulated by the lognormal or piecewise distribution. We further confirmed the deviated estimation through of survival curves in Figure 2, where we see the Weibull and exponential model can not provide the a fitted curve to match the observed. For these simulations, the Cox model provided small MSE values, and the fitted curve of Cox PH model was quite consistent with the observed data. Thus we demonstrated that Cox model can estimate the treatment effect with high accuracy and precision compared to a misspecified distribution, as expected.

On the other hand, in the case of the impact of fitting too complicated a model when an simpler model is sufficient, we found that there were minimal consequenses of the more complex models, in terms of MSE. For example, when fitting models on data from an exponential distribution, all three had very similar levels of

MSE. There was no case where the Cox model performed significantly worse than other models, in terms of MSE. This confirms the finding that the Cox proportional hazard model is semi-parametrically efficient, as even when a simpler model would suffice it still performs well. The Cox model's good perfomance in these cases is further demonstrated in Figure 2, where we see it follows the observed data closely, regardless of the underlying hazard curve the data was sampled from.

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