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 implement their respective 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 functions 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, which represents ~40% hazard reduction among those treated compared to the control group. 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 curves 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.

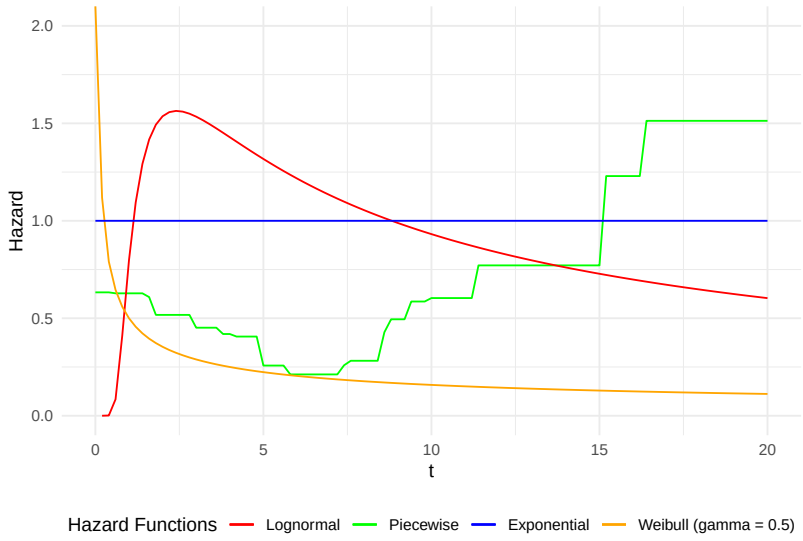

Figure 1. Baseline hazard functions used in this study for data generation. The lognormal distribution uses $\mu = 0$ and $\sigma = 0.5$. The piecewise distribution is based on the curve generated by Brilleman et al. The exponential and Weibull($\gamma = 0.5$) curves were fit using $\lambda = 1$. The 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 ($\hat{\beta}$). Furthermore, we generated six additional datasets of 100 samples from each baseline hazard distribution and fit exponential, Weibull, and Cox models for visualization in Figure 2 below.

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) and the mean and standard deviation of the estimated treatment effect among all 1000 simulations to evaluate model performance (Tables 2, 3).

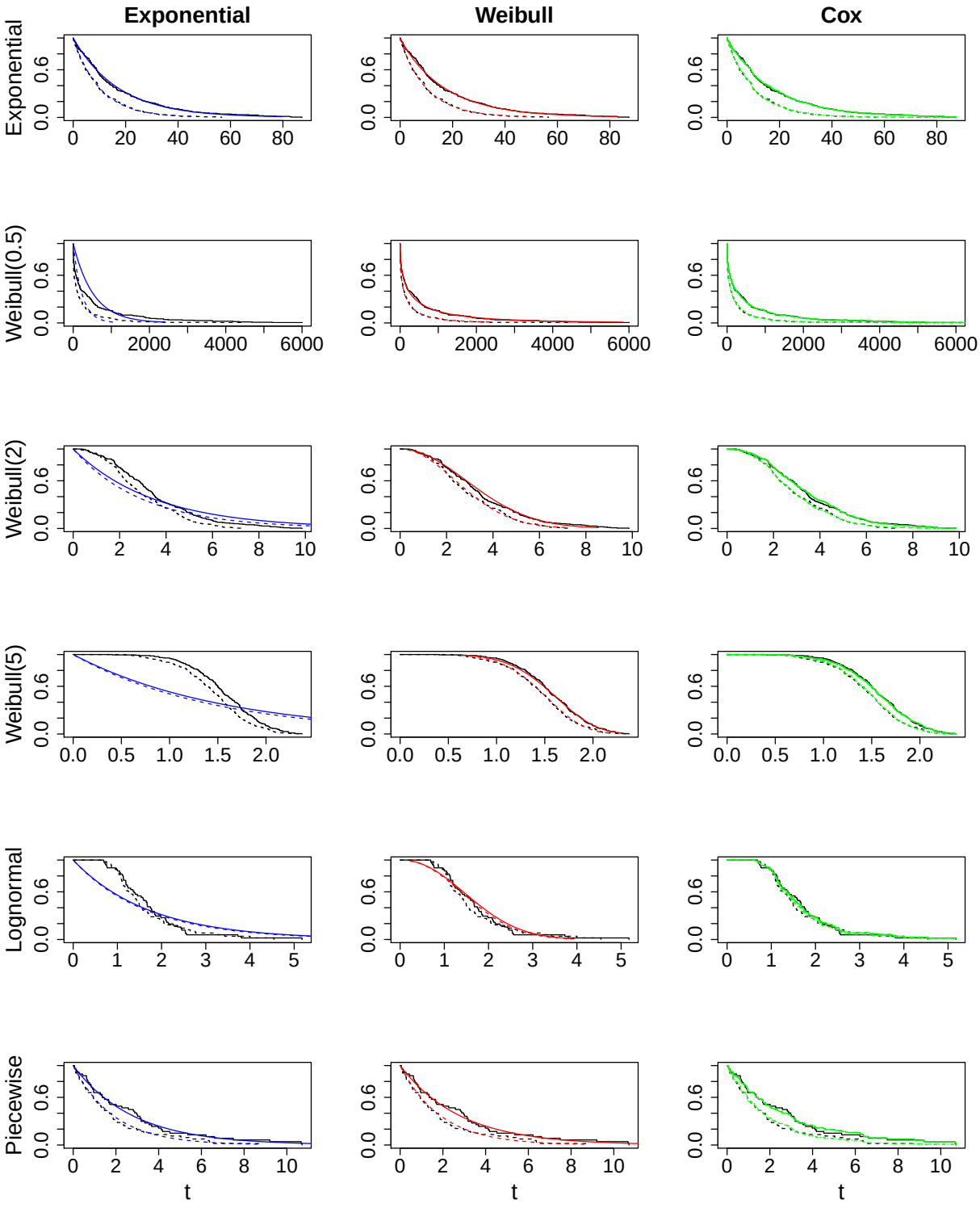

Figure 2. Survival probabilities versus time for all models (columns) and baseline hazard functions (rows). Solid and dashed lines represent the survival curves for the treatment and control groups, respectively. Kaplan-Meier curves are provided in black and model fits are plotted in blue (exponential), red (Weibull), and green (Cox).

In Figure 2 above, we can observe the impact of baseline hazard misspecification when fitting a proportional hazards model. For example, the exponential model cannot accurately fit a baseline hazard distribution of Weibull($\gamma = 2, 5$) due to concavity. For data simulated from the lognormal and piecewise baseline hazard functions, neither the exponential nor the Weibull models fit the survival data well. We also note that the Cox proportional hazards model has good fit across all six datasets, even when a simpler model would suffice. For instance, the Cox model appears to fit the simulated exponential data well, but the exponential model fits the data equally well. Thus, it may be more advantageous to utilize the exponential model in this case to increase parsimony and interpretability, and minimize model complexity.

In Table 2 below, we calculated the MSE of the treatment effect estimates from the three models and six baseline hazard functions. The Weibull model had the lowest MSE for data simulated from a Weibull hazard curve with $\gamma = 5$. The Cox proportional hazards model achieves the lowest MSE among data simulated from the exponential, Weibull with $\gamma = 0.5$, lognormal, and piecewise hazard curves. Interestingly, the exponential model performed best with the Weibull($\gamma = 2$) data, and its performance was nearly identical to the Cox model for data simulated from an exponential distribution. Similarly, for data simulated from Weibull($\gamma = 0.5$), the Cox model performed only slightly better than the Weibull model.

Model	Exponential	Weibull(0.5)	Weibull(2)	Weibull(5)	Lognormal	Piecewise
Exponential	0.1500	0.3298	0.0464	0.1615	0.0728	0.0558
Weibull	0.1511	0.3135	0.0488	0.0436	0.0936	0.0465
Cox	0.1489	0.3092	0.0489	0.0461	0.0423	0.0423

Table 2. MSEs from three models using 1000 simulations. Weibull base hazard functions are denoted as Weibull(γ).

Model	Exponential	Weibull(0.5)	Weibull(2)	Weibull(5)	Lognormal	Piecewise
Exponential	-0.523 (0.387)	-0.547 (0.573)	-0.350 (0.155)	-0.101 (0.045)	-0.248 (0.097)	-0.576 (0.224)
Weibull	-0.524 (0.388)	-0.533 (0.559)	-0.504 (0.221)	-0.509 (0.209)	-0.656 (0.263)	-0.528 (0.214)
Cox	-0.520 (0.386)	-0.529 (0.556)	-0.501 (0.221)	-0.507 (0.215)	-0.510 (0.205)	-0.510 (0.205)

Table 3. Mean (SD) of $\hat{\beta}$ from three models using 1000 simulations. Weibull base hazard functions are denoted as Weibull(γ).

In Table 3 above, we provide the breakdown of the average estimated treatment effect for each model and baseline hazard function. As the true treatment effect was $\beta = -0.5$, we see that the mean estimated treatment effect across all simulations was closest to this value for the Cox model, regardless of the baseline hazard function. We can see that in the case of the data simulated from the Weibull($\gamma = 5$) distribution and the lognormal distribution, fitting an exponential model introduces large bias in the estimation of the treatment effect, with mean estimates of -0.101 and -0.248, respectively. As far as the standard deviation of these estimates across the simulations, there is little difference between applying models that correctly specify the baseline hazard function and the Cox model. Interestingly, in the cases where the exponential model has been misspecified, the variation across samples of the estimates is small. This might explain why, in some instances, the exponential model outperforms the Weibull in terms of MSE, even when fit on Weibull simulated data, such as in the case of Weibull($\gamma = 2$).

Conclusions

By comparing the mean-squared error (MSE) and the mean and standard deviation of $\hat{\beta}$ obtained from the three models using 1000 simulations, we found that misspecifying the baseline hazard function can lead to significant differences between estimated treatment effect ($\hat{\beta}$) and the truth ($\beta = -0.5$).

For example, the exponential and Weibull models obtained relatively high MSEs due to biased means of $\hat{\beta}$ in data simulated from the lognormal and piecewise distributions. We further confirmed this deviated estimation through the survival curves provided in Figure 2, where we see the exponential and Weibull models cannot fit the simulated data

well due to distributional constraints. For these simulations, the Cox model provided the smallest MSE values, and the fitted curve was quite consistent with the simulated data. Thus, we have demonstrated that the Cox model can estimate the treatment effect with the highest accuracy and precision compared to a misspecified model, as expected.

Regarding the impact of fitting too complicated a model when an simpler model is sufficient, we found that there were minimal consequences of utilizing a more complex model in terms of MSE. For example, when fitting the proportional hazards models on data simulated from an exponential distribution, all three models had very similar MSEs, mean predicted values ($\hat{\beta}$), and respective standard deviations. There were no cases in which the Cox model performed significantly worse than the other models in terms of MSE; it outperformed other models in terms of bias. This confirms our finding that the semi-parametric Cox proportional hazards model is most efficient, as it is sufficient even when a simpler model would suffice. The appreciable performance of the Cox model is demonstrated in Figure 2, where we see that it follows the observed data closely, regardless of the underlying baseline hazard function that is specified.

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 and plots, 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