A simulation study comparing three survival models

Margaret Gacheru, Jin Ge, Sibei Liu & Adina Zhang

February 12, 2020

Project 1: Design a simulation study to compare three survival models

1. Evaluate the impacts of ***misspecifying*** the baseline hazard function on the estimate of the **treatment effect** $h(t|x) = h_0(t)e^{x^T\beta}$ $\frac{h(t|x_1)}{h(t|x_2)} = \exp[\beta^T(x_1 - x_2)]$
   a. Can we avoid this issue by using a semi-parametric Cox model?
   b. What might happen if we fit an over complicated model when exponential is sufficient?

Three survival models: 1) Proportional hazards $h_0(t) = \lambda$
2) Weibull proportional-hazards $h_0(t) = \lambda\gamma t^{\gamma-1}$
3) Cox proportional-hazards $h_0(t)$ unspecified.

Framework

Scenario 1: Assume **exponential** proportional baseline hazard

Scenario 2: Assume **Weibull** proportional baseline hazard

Scenario 3: Assume **Gompertz** proportional baseline hazard (unspecified hazard)

Step 1: Simulate survival data for each of the 3 scenarios

Step 2: Fit three survival models under each scenario

Step 3: Assess scenario performance (uncensored and censored data)

Step 4: Examine scenarios under different parameters (beta, sample size)

Simulation Data

Need to generate:

- **Treatment assignment**: treatment or control
- **Time-to-event** based on the baseline hazard function

In the case where data is censored (loss to follow-up, death unrelated to treatment, etc.):

- **Censored time-to-event**
- **Status** indicating whether individual is censored or not

Generating Time to Event

Utilize the relationship between survival and hazard baseline function

$$h(t|x) = h_0(t)e^{x^T\beta}$$

$$Given\ S(t|x) = e^{-H(t)}\ ,\ then\ S(t|x) = e^{-H_0(t)e^{x^T\beta}}$$

Use **Inverse Transformation Method** to obtain T

$$U = S(t|x) = e^{-H_0(t)e^{x^T\beta}}$$

$$T = H_0^{-1}\left(\frac{-log(U)}{e^{x^T\beta}}\right),\ U \sim U(0,1)$$

Characteristic	Distribution: Exponential	Weibull	Gompertz
Parameter	Scale parameter $\lambda > 0$	Scale parameter $\lambda > 0$ Shape parameter $\nu > 0$	Scale parameter $\lambda > 0$ Shape parameter $\alpha \in (-\infty, \infty)$
Range	$[0, \infty)$	$[0, \infty)$	$[0, \infty)$
Hazard function	$h_0(t) = \lambda$	$h_0(t) = \lambda \nu t^{\nu - 1}$	$h_0(t) = \exp(\alpha t)$
Cumulative hazard function	$H_0(t) = \lambda t$	$H_0(t) = \lambda t^{\nu}$	$H_0(t) = \frac{\lambda}{\alpha}(\exp(\alpha t) - 1)$
Inverse cumulative hazard function	$H_0^{-1}(t) = \lambda^{-1} t$	$H_0^{-1}(t) = (\lambda^{-1} t)^{1/\nu}$	$H_0^{-1}(t) = \frac{1}{\alpha} \log(\frac{\alpha}{\lambda} t + 1)$
Survival time	$T = -\frac{\log(U)}{\lambda \exp(\beta' x)}$	$T = (-\frac{\log(U)}{\lambda \exp(\beta' x)})^{1/\nu}$	$T = \frac{1}{\alpha} \log[1 - \frac{\alpha \log(U)}{\lambda \exp(\beta' x)}]$

Putting it all together

Treatment assignment:

$$X \sim Bernoulli(0.5)$$

Actual time to event:

$$T = H_0^{-1}\left(\frac{-log(U)}{e^{x^T\beta}}\right)$$

When data is censored:

$$Censored\ time \sim exponential(\lambda)$$

$$Status = \begin{cases} 1, & time\ to\ event < censored\ time \\ 0, & time\ to\ event > censored\ time \end{cases}$$

$$Observed\ time = min(time\ to\ event, censored\ time)$$

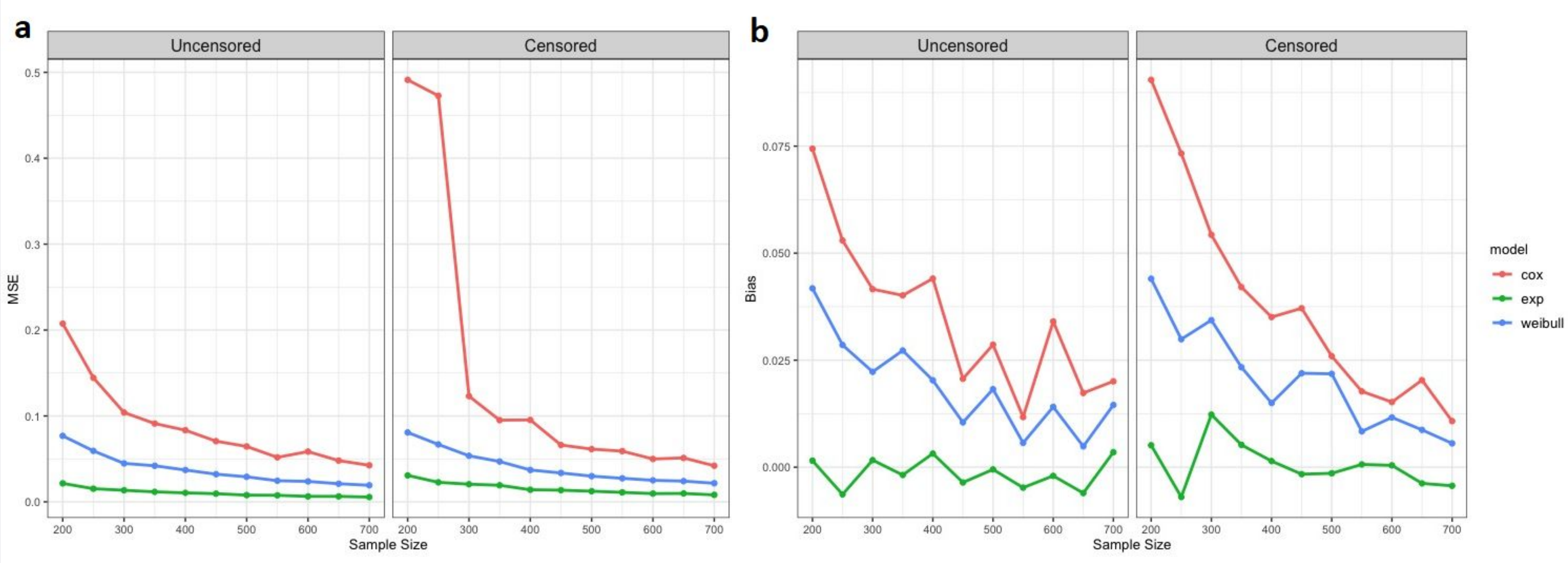
Point Estimates and 95% Confidence Intervals

Table 1. Beta point estimates and 95% confidence intervals across three scenarios

Model	Scenario 1 Exponential	Scenario 2 Weibull	Scenario 3 Gompertz
Exponential	4.00 (3.81, 4.19)	1.00 (0.95, 1.05)	1.90 (1.79, 2.02)
Weibull	4.03 (3.65, 4.40)	4.02 (3,66, 4.38)	3.11 (2.78, 3.44)
Cox	4.03 (3.46, 4.60)	4.04 (3.51, 4.56)	4.03 (3.49, 4.58)

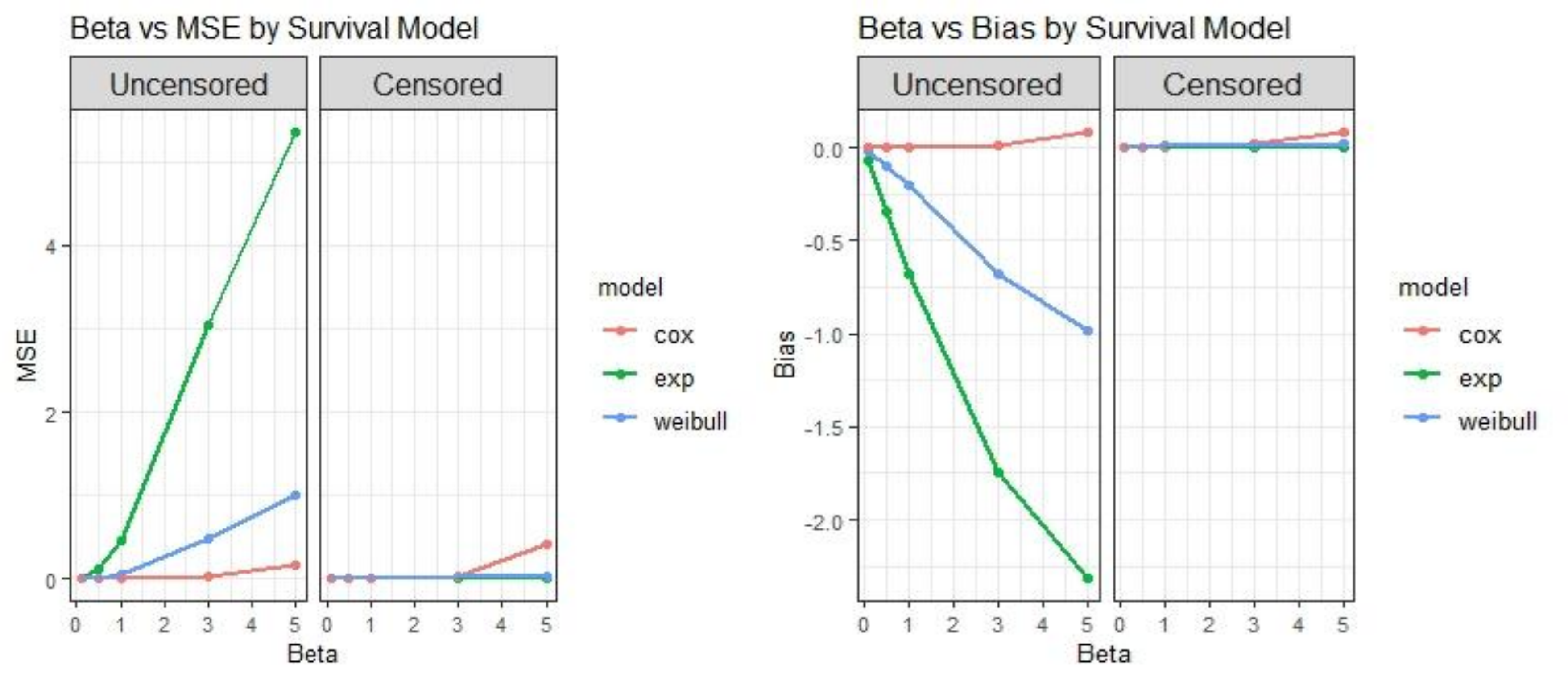
* Simulations = 1000, N = 400, beta = 4, lambda = 0.1, gamma = 4, alpha = 4, Uncensored

Scenario 1: Exponential-proportional hazards model



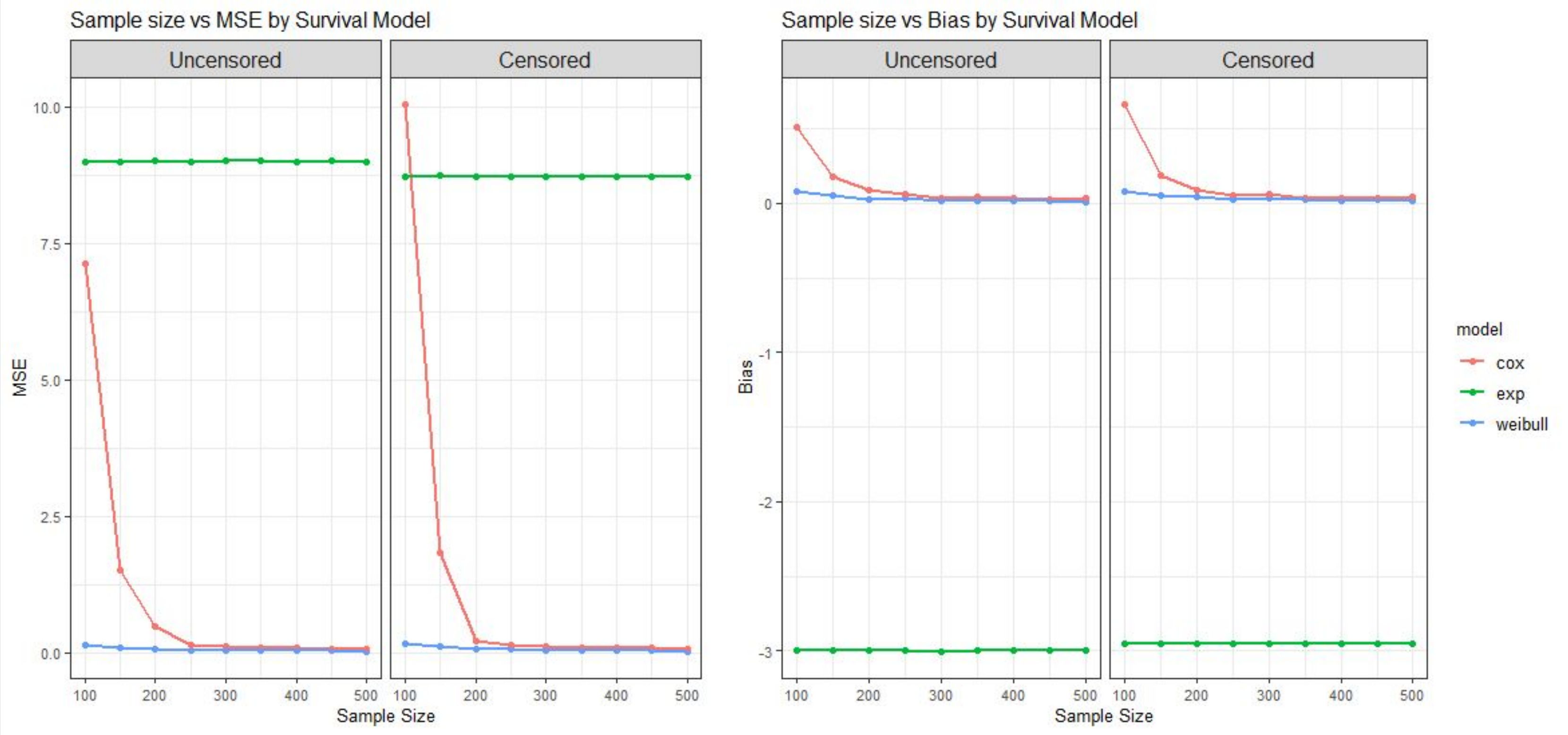
* Simulations = 1000, beta = 4, lambda = 0.1

Scenario 1 continued



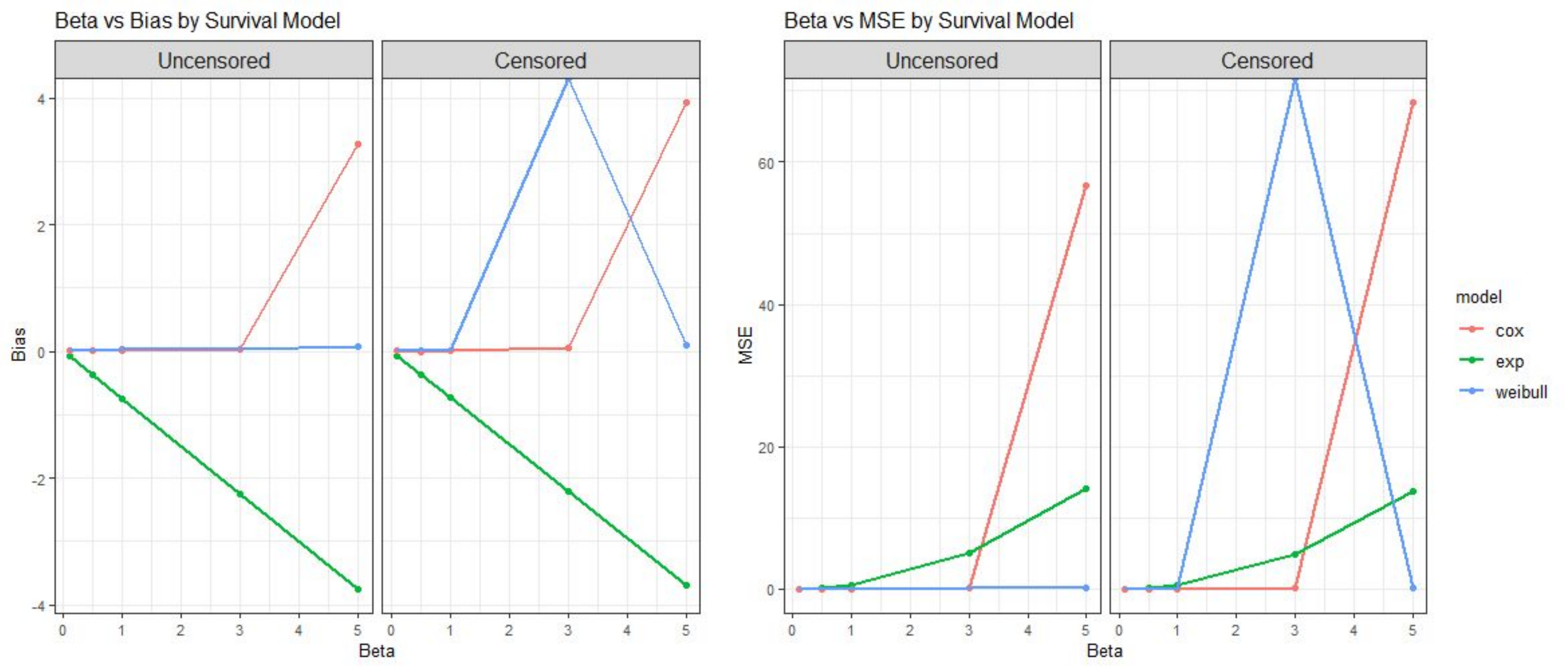
* Simulations = 1000, N = 500, lambda = 0.1

Scenario 2: Weibull-proportional hazards model



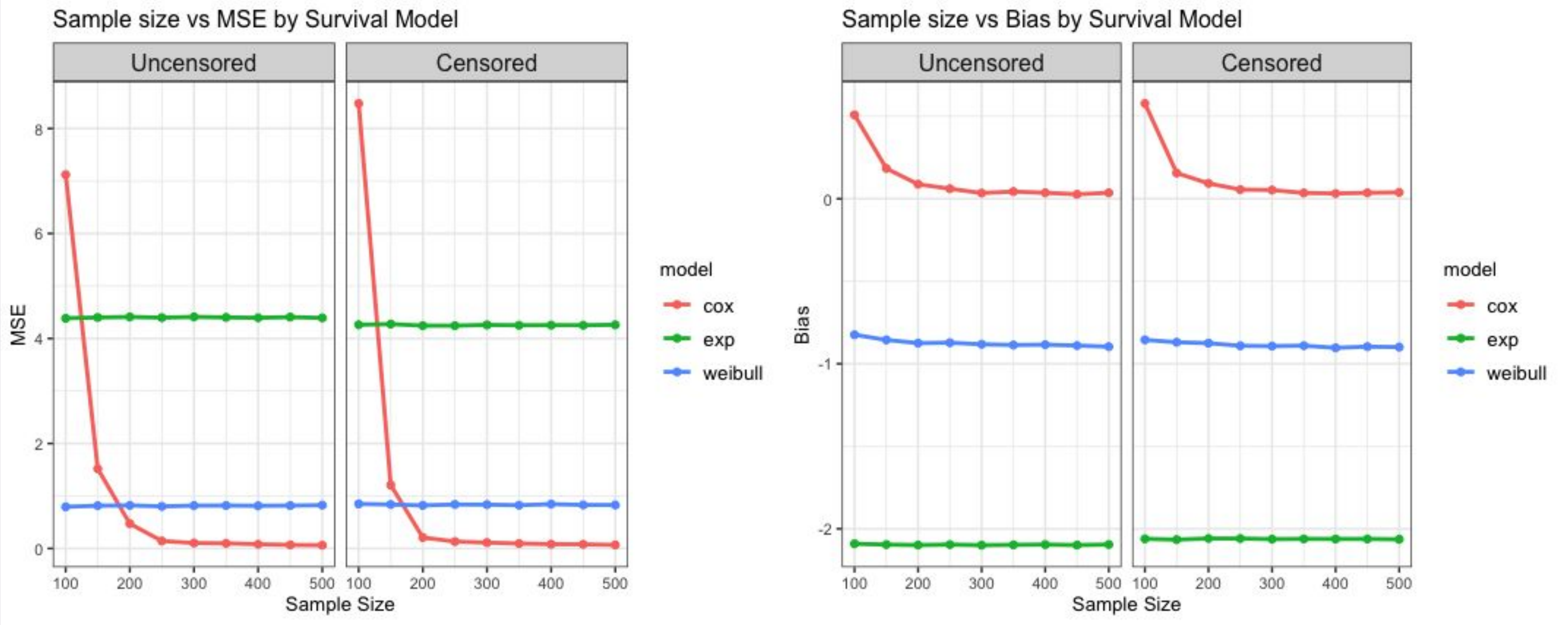
* Simulations = 1000, beta = 4, lambda = 0.1, gamma = 4, alpha = 4

Scenario 2 continued



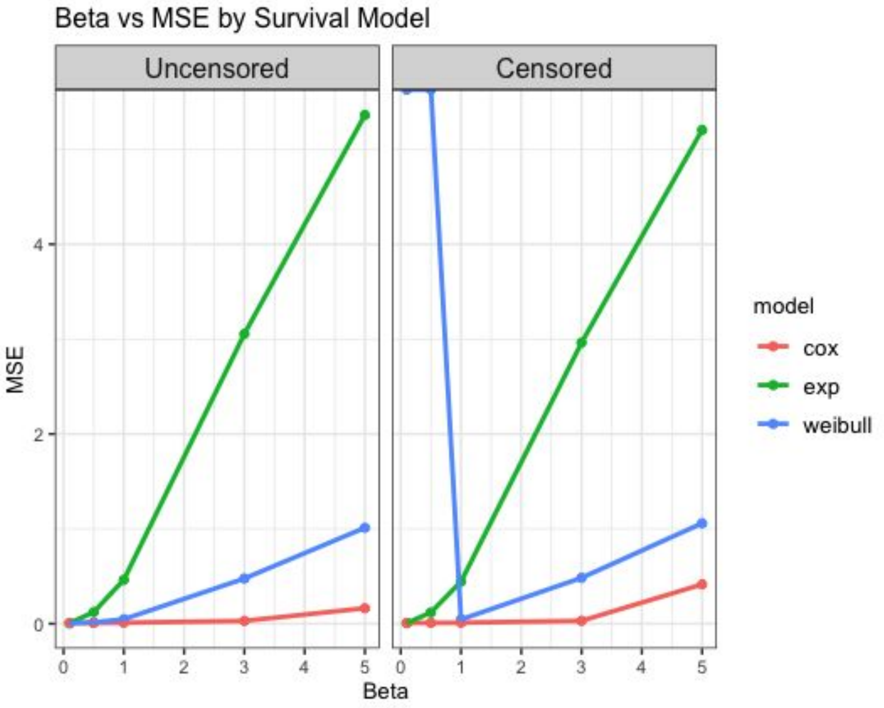
* Simulations = 1000, beta = 4, lambda = 0.1, gamma = 4, alpha = 4

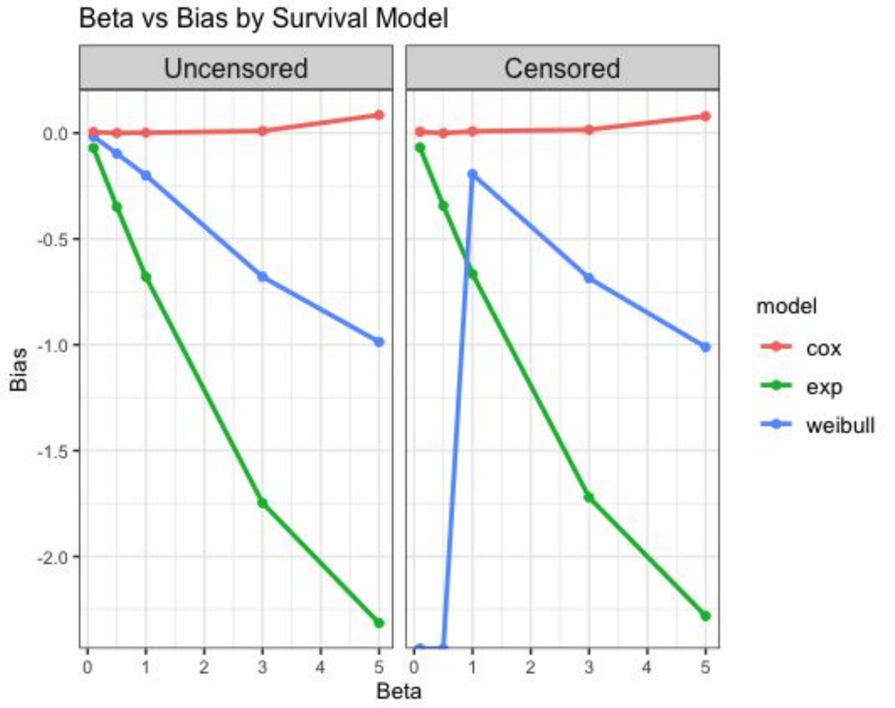
Scenario 3: Cox-proportional hazards model



* Simulations = 1000, beta = 4, lambda = 0.1, gamma = 4, alpha = 4

Scenario 3 continued





* Simulations = 1000, N = 500, lambda = 0.1, gamma = 4, alpha = 4

Conclusion

1. As expected, using the appropriate baseline hazard results in the best performance
2. When misspecifying a baseline hazard, it is best to use Cox
3. Mixed results when varying the treatment effect (possible convergence issues)

Further Exploration

- Varied hazard function inputs (ie. alpha, lambda, gamma)
- Inclusion of multiple covariates (continuous, categorical)
- Different censor time distributions
- Strong censoring vs weak censoring