Data Generation Plan for Simulation study

Neither **FAST** (Do et al., 2022) nor **FISA** (Keya et al., 2023) used simulated datasets in their experiments. They evaluated fairness-aware survival models on real medical data by manipulating follow-up times to mimic censoring disparity rather than generating synthetic event times from scratch. But to produce a simulation study for comparison across methods, I will design a synthetic data generator that reproduces the same bias mechanisms, unequal follow-up and censoring.

Data Generation Overview.

- A binary sensitive attribute like age, $A \in \{0, 1\}$ (e.g., two demographic groups with equal proportion).
- One or more covariates X (e.g., age, blood pressure) sampled from normal distributions to represent vital data.
- A true event time T generated from a Weibull distribution, chosen because it flexibly models increasing or decreasing risk over time while remaining easy to sample.
- A censoring time C generated independently from another Weibull or uniform distribution. To create group imbalance, group $A = 0$ will have a longer maximum observation window, while group $A = 1$ will have a shorter one.
- The observed time and event indicator, computed as $Y = \min(T, C, C_{\max}(A))$ and $\Delta = I\{T \leq \min(C, C_{\max}(A))\}$.

I will repeat the entire data generation and model fitting process across R Monte Carlo replications (e.g., $R = 100$), computing the mean and standard deviation of key metrics (C-index, Brier score, and group disparities).

This setup creates a clear, tunable difference in censoring across groups. It allows testing across methods like FAST (which enforces independence between predicted time-to-event and A) and FISA (which constrains hazard-level fairness).

Next Step. Generate data using R. And begin with FAST, test its performance and compare with baseline models.