Supplementary Material "A Bayesian Nonparametric Approach for Evaluating the Causal Effect of Treatment in Randomized Trials with Semi-Competing Risks"

Yanxun Xu, Daniel Scharfstein, Peter Müller, Michael Daniels

1 Data Analysis for a Simulated Dataset

Since the data from the brain tumor study used in the paper come from a clinical trial and cannot be made available online, we simulated a dataset with similar design as the real data to illustrate how the code can be used to analyze this simulated dataset.

We first applied the proposed BNP model to the simulated dataset. An initial analysis of the brain tumor death outcome using Kaplan-Meier is given in Figure S1, indicating that the treatment group has higher estimated survival probabilities. Figure S1 plots the estimated posterior survival curves for treatment and control groups marginalized over the distribution of covariate with 95% credible intervals under the proposed BNP method.

Figure S2 plots the posterior estimates (along with point-wise 95% credible intervals) of the causal estimand $\tau(u)$ versus u for two choices of ρ, 0.2, and 0.8. The estimand $\tau(u)$ contrasts the risk of progression prior to time u for treatment relative to control. If

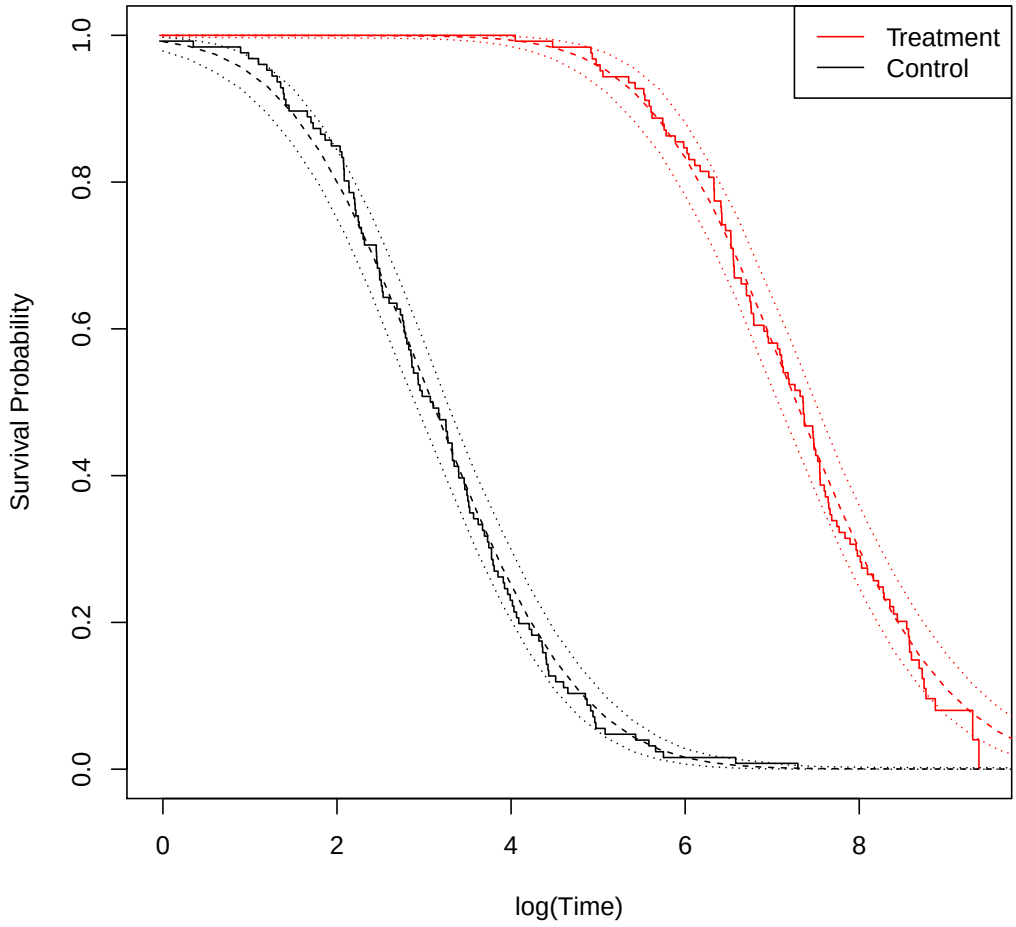

Figure S1: The dashed lines represent the estimated posterior mean survival curves for the proposed BNP method. The solid lines represent the Kaplan-Meier curves of the observed survival data in control and treatment groups, and the dotted lines represent 95% point-wise credible intervals of the posterior estimated survival curves. Survival times are on the log scale (days).

$\tau(u) < 1$, then among the patients who would survive to time u under both treatments, the probability of experiencing neurological progression prior to time u is lower under treatment than under control. As seen in Figure S2, when $\rho = 0.8$, the probability of experiencing neurological progression prior to time u is lower under treatment than under control. When $\rho = 0.2$, we still get the same conclusion based on the posterior estimated $\tau(u)$, but there is appreciable uncertainty characterized by wide posterior

credible intervals when $u > 0.45$.

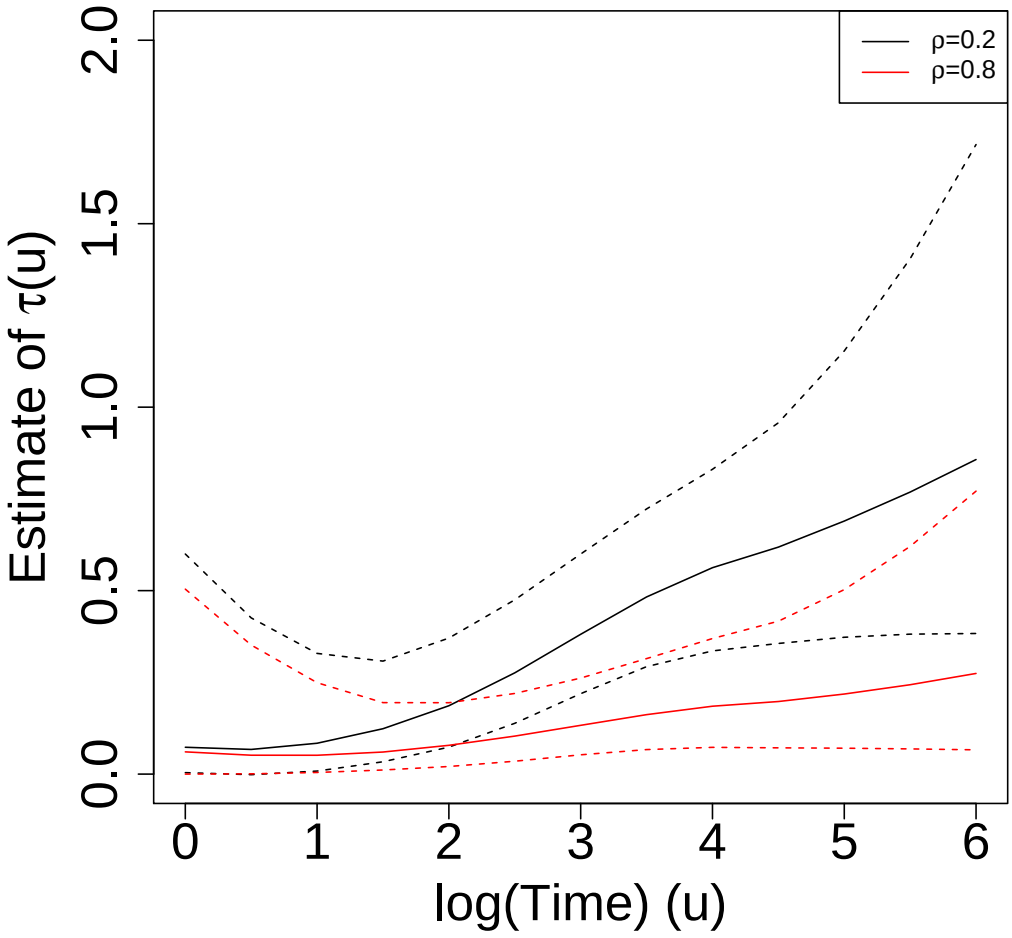

Figure S2: Posterior estimated $\tau(u)$ versus u on the log scale (days) in the simulated dataset for different ρ's under the proposed BNP method. The solid lines represent the posterior estimated $\tau(u)$, and the dashed lines represent 95% point-wise credible intervals.