Supplementary Material for "A Unified Three-State Model Framework for Analysis of Treatment Crossover in Survival Trials"

Zile Zhao[a], Ye Li[b], Xiaodong Luo[b], Ray Bai[a*]

[a]Department of Statistics, University of South Carolina, Columbia, SC 29208
[b]Evidence Generation & Decision Science, R&D, Sanofi US, Bridgewater, NJ 08807

November 20, 2024

Abstract

In Section A, we describe the random number generation process for generating crossover time and event time in the TSM framework. In Section B, we present additional simulation results under varying proportions of censored patients.

A Random Number Generation in the TSM Framework

Based on the TSM framework and the additional assumption $P(U > T) = 0$, we can write the conditional hazard function of T given $U = u$ as

$$\lambda(t \mid u) = \lambda_1(t)I(t \leq u) + \lambda_2^{\mathbf{x}}(t \mid u)I(t > u).$$

The univariate function $\lambda_1(t)$ is the hazard rate without crossover and the bivariate function $\lambda_2^{\mathbf{x}}(t \mid u)$ is the hazard rate from crossover to event. This second hazard rate may depend on when the crossover occurs. Note that the joint distribution can be viewed as being conditional on $U \leq T$, and T and U are not assumed to be independent of each other.

*Corresponding author: rbai@mailbox.sc.edu. The authors gratefully acknowledge financial support from the McCausland Innovation Fund and the University of South Carolina Office of the Vice President for Research ASPIRE program.

Markov crossover. When $\lambda_2^{\mathbf{x}}(t \mid u) = \lambda_2(t)$, the survival function $\widetilde{S}(t)$ of the event time T is

$$\widetilde{S}(t) = P(T > t) = S_1(t)S_3(t) + \int_0^t \frac{S_1(u)S_3(u)S_2(t)}{S_2(u)}\lambda_3(u)du,$$

where $S_k(t) = \exp\{-\int_0^t \lambda_k(u)du\}$, $k = 1, 2, 3$. To generate (T, U), we use the following steps:

(MC1) Generate two independent random variables X and U such that X follows Unif$(0, 1)$ and U follows $F_3 = 1 - S_3$.

(MC2) If $X > S_1(U)$, set $T = F_1^{-1}(1 - X)$; else set $T = F_2^{-1}\{1 - XS_2(U)/S_1(U)\}$, where F_k^{-1} is the inverse function of $F_k = 1 - S_k$, $k = 1, 2$.

Semi-Markov crossover. When $\lambda_2^{\mathbf{x}}(t \mid u) = \lambda_2(t - u)$, the marginal survival function $\widetilde{S}(t)$ of the event time is

$$\widetilde{S}(t) = P(T > t) = S_1(t)S_3(t) + \int_0^t S_1(u)S_3(u)S_2(t - u)\lambda_3(u)du,$$

where $S_k(t) = \exp\{-\int_0^t \lambda_k(u)du\}$, $k = 1, 2, 3$. To generate (T, U), we use the following steps:

(SMC1) Generate two independent random variables X and U such that X follows Unif$(0, 1)$ and U follows $F_3 = 1 - S_3$.

(SMC2) If $X > S_1(U)$, set $T = F_1^{-1}(1 - X)$; else set $T = U + F_2^{-1}\{1 - X/S_1(U)\}$, where F_k^{-1} is the inverse function of $F_k = 1 - S_k$, $k = 1, 2$.

General crossover. For general $\lambda_2^{\mathbf{x}}(t \mid u)$, the marginal survival function,

$$\widetilde{S}(t) = P(T > t) = S_1(t)S_3(t) + \int_0^t S_1(u)S_3(u)S_2(t \mid u)\lambda_3(u)du,$$

where $S_2(t \mid u) = \exp\{-\int_u^t \lambda_2^{\mathbf{x}}(s \mid u)ds\}$. We use the following steps:

(GC1) Generate two independent random variables X and U such that X follows Unif$(0, 1)$ and U follows $F_3 = 1 - S_3$.

(GC2) If $X > S_1(U)$, set $T = F_1^{-1}(1 - X)$, where F_1^{-1} is the inverse function of $F_1 = 1 - S_1$.

(GC3) If $X \leq S_1(U)$, set

$$T = F_{2,U}^{-1}\{1 - X/S_1(U)\},$$

where, $S_{2,u}(t) = S_2(t \mid u)$, $F_{2,u} = 1 - S_{2,u}$ and $F_{2,u}^{-1}$ is the inverse function of $F_{2,u}$ for any fixed u.

This approach to generating the crossover time and event time in TSM (assuming that all the pertaining hazard functions are piecewise constant functions) has been implemented using the function `rpwecx` in the R package `PWEALL` (version 1.4.0 on CRAN).

B Additional Simulation Results

In this section, we examine the effect of the censoring rate on the different TSM methods. We repeated the experiments from Section 4.1 under lower and higher censoring rates than those of the simulations in Section 4.1. Our experiments in Section 4.1 had approximately 38.4%, 40%, 41.6% and 43.2% censored patients respectively when $\pi_2 = 0.25$, 0.5, 0.75 and 1. We deem these scenarios as "moderate censoring."

To attain lower censoring rates, we assumed that the yearly censoring rate was 0.02. The study had a one-year recruitment period with a uniform accrual rate to recruit 400 patients, and the study would be read out at 8 years after the randomization of the first patient. This resulted in roughly 28.4%, 29.9%, 31.4% and 32.9% censored patients respectively when $\pi_2 = 0.25$, 0.5, 0.75 and 1.

To attain higher censoring rates, we assumed that the yearly censoring rate was 0.025. The study had a two-year recruitment period with a uniform accrual rate to recruit 400 patients, and the study would be read out at 5 years after the randomization of the first patient. This resulted in roughly 50.5%, 51.9%, 53.2% and 54.6% censored patients respectively when $\pi_2 = 0.25$, 0.5, 0.75 and 1.

The results for lower censoring are shown in Figure 1 and Table 1, while the results for higher censoring are shown in Figure 2 and Table 2. Overall, the average bias and SE for all methods tended to be slightly lower under lower censoring than moderate censoring. Meanwhile, average bias and MSE tended to be higher under higher censoring. This is what we expect since higher censoring percentages inherently make estimation of the true treatment effect more difficult. Interestingly, however, the censoring percentage did not have as much of an impact on the ECP. This could be because higher amounts of censoring also tended to lead to higher SE's (or greater uncertainty) and therefore more conservative 95% CIs.

Regardless of lower, moderate, or higher censoring rates, our results in Figures 1 and 2 and Tables 1 and 2 are consistent with those in Section 4.1.2. Namely, the TSM methods that assumed

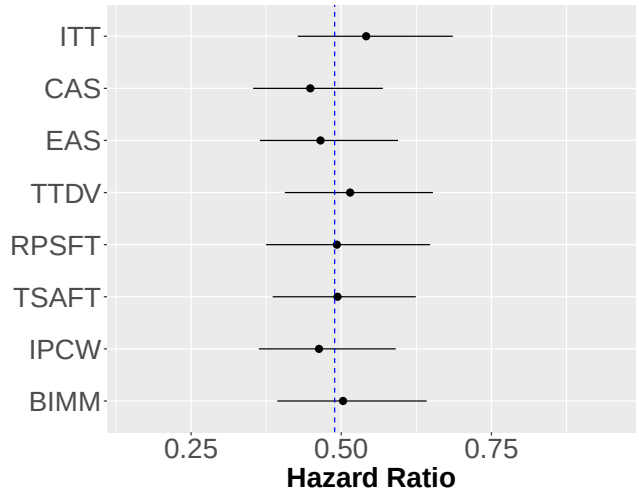
semi-Markov crossover and that used the data after crossover (i.e. RPSFT, TSAFT, and BIMM) consistently outperformed the other methods (ITT, CAS, EAS, TTDV, and IPWC). Moreover, the performance of ITT, CAS, EAS, TTDV, and IPWC degraded considerably as the crossover propotion π_2 increased. This suggests that under many different scenarios with different censoring and crossover proportions, it may be prudent for practitioners to consider using the data after crossover to estimate treatment effects.

Table 1: Simulation results under lower censoring based on 2000 replications. The table displays the average Bias, SE, and MSE from all Monte Carlo replicates, while the ECP is the percentage of 95% CIs that contained the true HR.

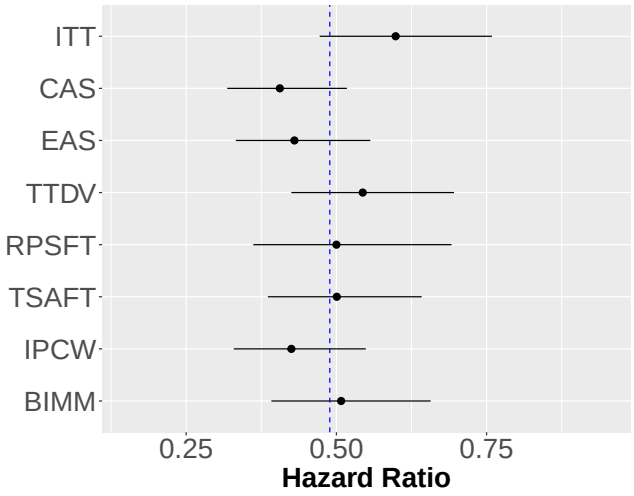
	25% crossover				**50% crossover**			
Method	Bias	SE	MSE	ECP (%)	Bias	SE	MSE	ECP (%)
ITT	0.052	0.065	0.011	87.1	0.109	0.072	0.023	63.9
CAS	-0.040	0.054	0.007	87.6	-0.083	0.050	0.012	65.4
EAS	-0.024	0.058	0.007	92.2	-0.058	0.056	0.010	82.7
TTDV	0.025	0.062	0.008	92.9	0.054	0.063	0.012	87.3
RPSFT	0.003	0.075	0.010	96.0	0.012	0.090	0.014	96.0
TSAFT	0.004	0.062	0.008	93.3	0.011	0.066	0.009	93.3
IPCW	-0.025	0.057	0.007	92.2	-0.063	0.055	0.010	81.0
BIMM	0.014	0.062	0.008	94.0	0.018	0.066	0.009	94.3
	75% crossover				**100% crossover**			
Method	Bias	SE	MSE	ECP (%)	Bias	SE	MSE	ECP (%)
ITT	0.164	0.079	0.039	36.9	0.222	0.087	0.065	16.2
CAS	-0.135	0.045	0.022	26.2	-0.196	0.039	0.042	1.9
EAS	-0.119	0.052	0.020	47.4	-0.235	0.041	0.059	1.6
TTDV	0.092	0.078	0.021	76.6	0.157	0.100	0.046	60.1
RPSFT	0.014	0.109	0.021	97.5	0.022	0.140	0.034	98.9
TSAFT	0.016	0.078	0.012	94.0	0.017	0.125	0.031	93.5
IPCW	-0.128	0.051	0.021	39.3	-0.254	0.038	0.067	0.15
BIMM	0.015	0.074	0.011	95.0	0.013	0.070	0.019	77.1

Table 2: Simulation results under higher censoring based on 2000 replications. The table displays the average Bias, SE, and MSE from all Monte Carlo replicates, while the ECP is the percentage of 95% CIs that contained the true HR.

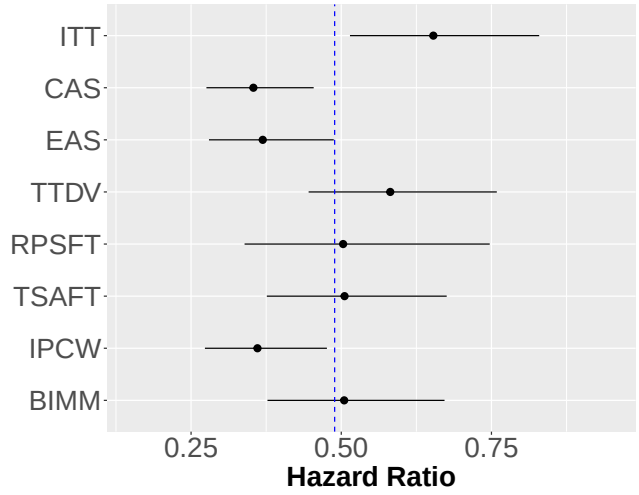
	25% crossover				**50% crossover**			
Method	Bias	SE	MSE	ECP (%)	Bias	SE	MSE	ECP (%)
ITT	0.042	0.086	0.017	91.8	0.084	0.094	0.025	83.9
CAS	-0.024	0.076	0.013	92.8	-0.054	0.072	0.014	87.7
EAS	-0.020	0.078	0.013	93.3	-0.052	0.076	0.015	87.6
TTDV	0.027	0.083	0.015	93.9	0.055	0.089	0.020	89.8
RPSFT	0.011	0.099	0.018	95.4	0.021	0.112	0.023	96.2
TSAFT	0.009	0.085	0.015	93.9	0.016	0.090	0.017	93.7
IPCW	-0.027	0.077	0.013	92.7	-0.070	0.072	0.015	83.7
BIMM	0.023	0.085	0.015	93.7	0.028	0.089	0.017	94.0
	75% crossover				**100% crossover**			
Method	Bias	SE	MSE	ECP (%)	Bias	SE	MSE	ECP (%)
ITT	0.129	0.102	0.038	73.2	0.178	0.111	0.056	56.4
CAS	-0.087	0.068	0.017	76.6	-0.126	0.064	0.024	59.0
EAS	-0.099	0.071	0.020	73.9	-0.172	0.063	0.038	37.4
TTDV	0.088	0.099	0.027	85.6	0.135	0.114	0.045	76.4
RPSFT	0.034	0.131	0.032	97.1	0.050	0.157	0.047	98.1
TSAFT	0.027	0.106	0.022	94.3	-0.045	0.148	0.046	94.0
IPCW	-0.130	0.065	0.025	55.4	-0.225	0.051	0.055	6.25
BIMM	0.029	0.095	0.019	95.0	0.045	0.094	0.030	82.8



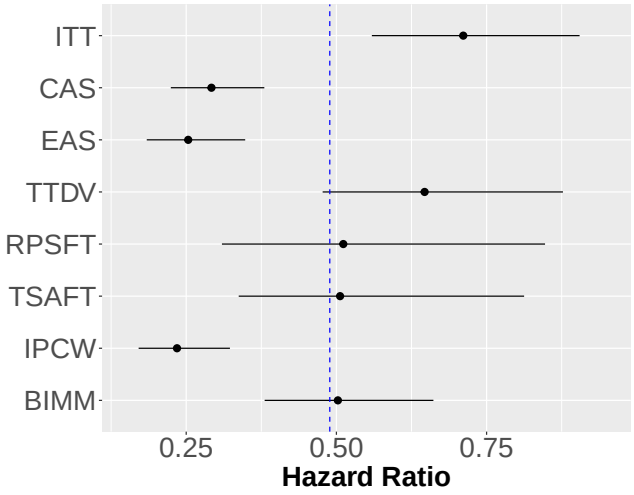
(a) 25% crossover after progression



(b) 50% crossover after progression

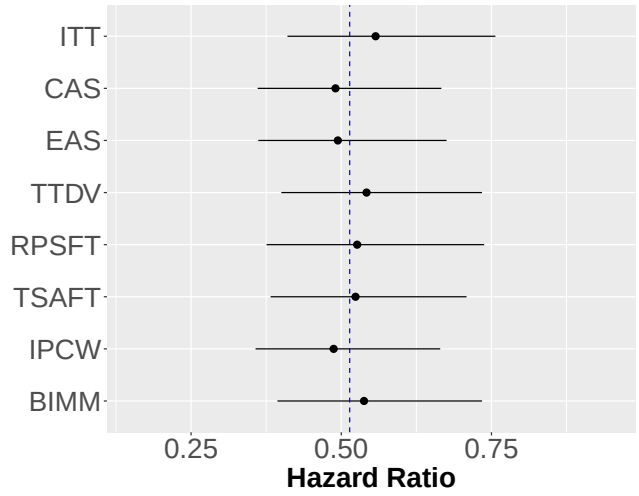


(c) 75% crossover after progression

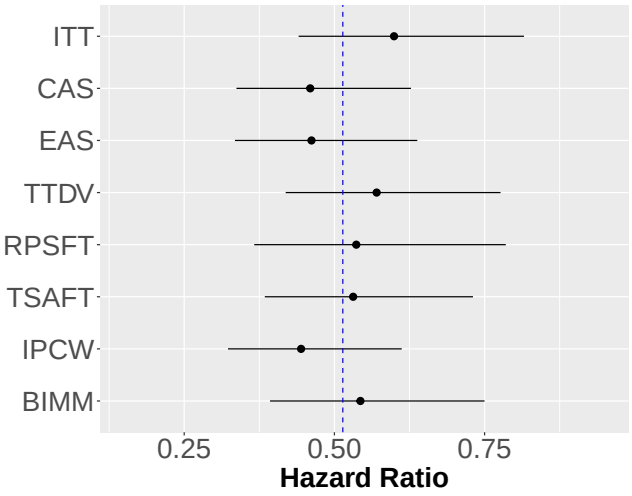


(d) 100% crossover after progression

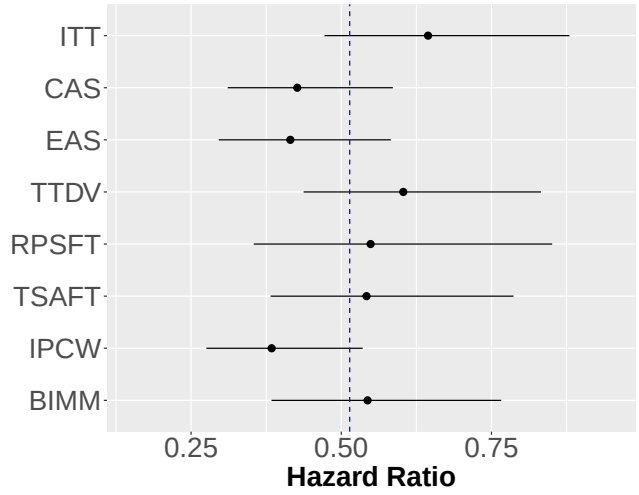
Figure 1: Simulation results under lower censoring averaged across 2000 replicates. The dotted blue line is the true treatment effect ($\text{HR}_{\text{true}} = 0.489$).



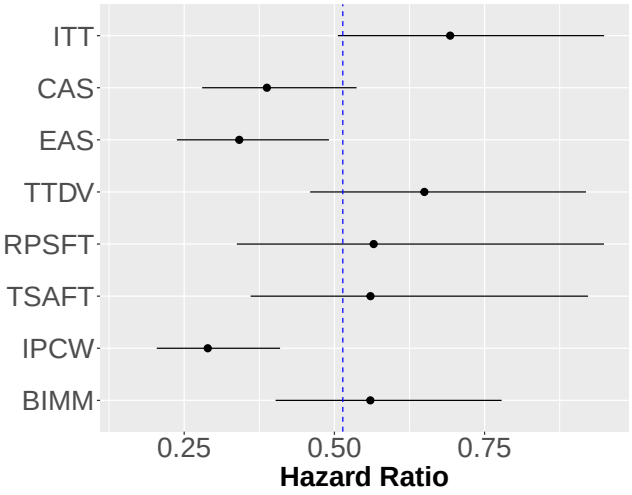
(a) 25% crossover after progression



(b) 50% crossover after progression



(c) 75% crossover after progression



(d) 100% crossover after progression

Figure 2: Simulation results under higher censoring averaged across 2000 replicates. The dotted blue line is the true treatment effect ($\text{HR}_{\text{true}} = 0.514$).