Web-Based Supplementary Materials for Bias Induced by Ignoring Double Truncation Inherent in Autopsy-Confirmed Survival Studies of Neurodegenerative Diseases by Lior Rennert and Sharon X. Xie

August 16, 2018

Web Appendix A

Here we justify that the identifiability constraints given in Section 2, $a_{H_L} < a_F \leq a_{H_R}$ and $b_{H_L} \leq b_F < b_{H_R}$, hold in our data example. First we introduce some notation. Denote τ as the end of study date, τ_E as the study entry date, and τ_A as the date of symptom onset. Note that τ is the same for all subjects, while τ_E and τ_A can differ among subjects. The left truncation time is defined as $L = \tau_E - \tau_A$, and the right truncation time is defined as $R = \tau - \tau_A$.

Subjects can theoretically enter the center at the time of AD or FTLD symptom onset, but not before. Therefore the smallest possible left truncation time is $L = 0$, and thus $a_{H_L} = 0$. Since recruitment of subjects into the center stops one week prior to the end of the study, the smallest possible right truncation time is $R = 1$ week (or $R \approx 0.019$ years). Therefore $a_{H_R} \geq 0.019$. Since subjects can die from AD or FTLD within a week of symptom onset, and we assume that subjects cannot die on the day of symptom onset, $P(T \leq t) > 0$ for some $t \in (0, 0.019)$, where t is measured in years. We therefore have that $0 < a_F < 0.019$, and thus the constraint $a_{H_L} < a_F \leq a_{H_R}$ is satisfied.

Because subjects are not expected to enter the study more than 15 years after symptom onset, the assumption $b_{H_L} \leq 15$ is reasonable in practice. Since subjects with AD or FTLD can live past 15 years after symptom onset, $P(T \geq 15) > 0$, and thus $b_F \geq 15$. Our study recruited subjects from 1995 to 2012, and subjects with AD or FTLD symptom onset before 1995 were included in the study. Since the study recruited subjects over a 17 year period, we have that $b_{H_R} = b_F + 17$ and thus $b_F < b_{H_R}$. To see why this is so, note that a subject with a survival time $T = b_F$ could have theoretically had symptom onset in the year 1995 - b_F. For example, if $b_F = 15$, a subject could have entered the study in 1980, in which case their right truncation time would be 32 years. Therefore it is reasonable to assume that the constraint $b_{H_L} \leq b_F < b_{H_R}$ is satisfied.

Notation for Web Tables 1,2,3, and 4

In the following tables, q_L, q_R, q are the proportion of observations missing due to left, right, and double (left and right) truncation, respectively, and n is the size of the observed sample. $\hat{F}_{SP}$ denotes the SPMLE, $\hat{F}_{NP}$ denotes the NPMLE, and $\hat{F}_{emp}$ denotes the naïve empirical CDF which ignores double truncation. These estimators were all computed at $t_{0.1}, ..., t_{0.9}$, the 1st through 9th deciles of the true survival distribution F_0. For a given estimator $\hat{F}$, Bias($\hat{F}$) is the (absolute) difference between $\hat{F}$ and F_0, averaged across the 9 deciles. The standard deviation of $\hat{F}$ across simulations is SD($\hat{F}$) is, $\widehat{\text{SE}}(\hat{F})$ is the estimated standard error of $\hat{F}$, MSE($\hat{F}$) is the mean squared error of $\hat{F}$, and Cov($\hat{F}$) is 95% coverage, all averaged across the 9 deciles. The estimated median value based on $\hat{F}$ is $\hat{t}_{0.5}$. The true median value based on F_0 is $t_{0.5} = 9.7$, Bias($\hat{t}_{0.5}$) = $\hat{t}_{0.5}$ - $t_{0.5}$, and SD($\hat{t}_{0.5}$) is the standard deviation of $\hat{t}_{0.5}$ across simulations.

Web Table 1

Table 1: Simulation results under misspecification of the truncation distribution

Survival times simulated from a $gamma(10, 1)$ distribution. Left and right truncation times assumed to come from a $gamma(\alpha_1, \beta_1)$ and $gamma(\alpha_2, \beta_2)$ distribution, respectively. Model A.1 corresponds to misspecification of the right truncation time by simulating it as $Unif(0, 20)$, and the left truncation time as $gamma(4.5, 1.5)$. Model A.2 corresponds to misspecification of the left truncation time by simulating it as $Weibull(3, 9)$, and the right truncation time as $gamma(5, 2)$. Model A.3 corresponds to misspecification both truncation times by simulating the left truncation time as $Weibull(3, 9)$ and the right truncation time as $Unif(0, 20)$.

Model	q_L, q_R, q	n	Estimator	Bias($\hat{F}$)	SD($\hat{F}$)	$\widehat{\text{SE}}(\hat{F})$	MSE($\hat{F}$)	Cov($\hat{F}$)	Bias($\hat{t}_{0.5}$)	SD($\hat{t}_{0.5}$)
			$\hat{F}_{SP}$	0.013	0.078	0.248	0.006	0.974	-0.060	0.729
A.1	0.22,0.50,0.64	50	$\hat{F}_{NP}$	0.025	0.080	0.073	0.007	0.872	-0.048	0.796
			$\hat{F}_{emp}$	0.037	0.056	0.053	0.005	0.484	-0.259	0.464
			$\hat{F}_{SP}$	0.013	0.034	0.083	0.001	0.997	-0.061	0.316
A.1	0.22,0.50,0.64	250	$\hat{F}_{NP}$	0.009	0.038	0.034	0.002	0.927	-0.050	0.442
			$\hat{F}_{emp}$	0.037	0.024	0.024	0.002	0.334	-0.225	0.205
			$\hat{F}_{SP}$	0.010	0.090	0.120	0.008	0.885	-0.008	0.990
A.2	0.33,0.52,0.74	50	$\hat{F}_{NP}$	0.023	0.098	0.086	0.010	0.863	0.018	1.079
			$\hat{F}_{emp}$	0.051	0.054	0.052	0.006	0.462	-0.292	0.424
			$\hat{F}_{SP}$	0.009	0.040	0.036	0.002	0.880	-0.056	0.342
A.2	0.33,0.52,0.74	250	$\hat{F}_{NP}$	0.006	0.042	0.041	0.002	0.928	-0.023	0.376
			$\hat{F}_{emp}$	0.051	0.025	0.023	0.004	0.328	-0.267	0.189
			$\hat{F}_{SP}$	0.010	0.087	0.317	0.008	0.967	-0.090	0.853
A.3	0.33,0.50,0.70	50	$\hat{F}_{NP}$	0.023	0.090	0.081	0.009	0.880	-0.057	0.936
			$\hat{F}_{emp}$	0.043	0.056	0.052	0.006	0.478	0.259	0.459
			$\hat{F}_{SP}$	0.012	0.037	0.098	0.002	0.995	-0.093	0.331
A.3	0.33,0.50,0.70	250	$\hat{F}_{NP}$	0.008	0.039	0.037	0.002	0.922	-0.037	0.341
			$\hat{F}_{emp}$	0.043	0.025	0.024	0.003	0.333	0.270	0.205

Web Table 2

Table 2: Simulation results when $R = L + d_0$.

Here d_0 is constant. Survival times simulated from a gamma$(10, 1)$ distribution. Left truncation time assumed to come from a gamma(α_1, β_1) distribution. Model B.1 corresponds to correct specification of left truncation time by simulating $L \sim$ gamma$(4.5, 1.5)$ and right truncation time $R \sim L + 6$. Model B.2 corresponds to misspecification of the left truncation time by simulating it as $Weibull(3, 9)$, and $R \sim L + 6$. Model B.3 corresponds to correct specification of left truncation time by simulating $L \sim$ gamma$(4.5, 1.5)$ and right truncation time $R \sim L + 8$. Model B.4 corresponds to misspecification of the left truncation time by simulating it as $Weibull(3, 9)$, and $R \sim L + 8$.

Model	d_0	q_L, q_R, q	n	Estimator	Bias($\hat{F}$)	SD($\hat{F}$)	$\widehat{\text{SE}}(\hat{F})$	MSE($\hat{F}$)	Cov($\hat{F}$)	Bias($\hat{t}_{0.5}$)	SD($\hat{t}_{0.5}$)
				$\hat{F}_{SP}$	0.008	0.105	0.070	0.011	0.787	0.069	1.167
B.1	6	0.22,0.26,0.48	50	$\hat{F}_{NP}$	0.026	0.106	0.093	0.012	0.824	0.070	1.158
				$\hat{F}_{emp}$	0.067	0.055	0.050	0.009	0.449	-0.498	0.408
				$\hat{F}_{SP}$	0.006	0.050	0.035	0.003	0.828	-0.016	0.479
B.1	6	0.22,0.26,0.48	250	$\hat{F}_{NP}$	0.008	0.051	0.050	0.003	0.918	-0.019	0.473
				$\hat{F}_{emp}$	0.067	0.024	0.023	0.007	0.297	-0.479	0.177
				$\hat{F}_{SP}$	0.009	0.098	0.069	0.010	0.789	0.000	0.920
B.2	6	0.33,0.17,0.50	50	$\hat{F}_{NP}$	0.025	0.105	0.099	0.012	0.868	-0.011	1.070
				$\hat{F}_{emp}$	0.055	0.053	0.051	0.007	0.456	0.331	0.418
				$\hat{F}_{SP}$	0.012	0.045	0.033	0.002	0.802	-0.046	0.387
B.2	6	0.33,0.17,0.50	250	$\hat{F}_{NP}$	0.006	0.050	0.049	0.003	0.931	-0.010	0.455
				$\hat{F}_{emp}$	0.056	0.024	0.023	0.005	0.327	0.351	0.192
				$\hat{F}_{SP}$	0.003	0.083	0.068	0.007	0.863	0.057	0.812
B.3	8	0.22,0.14,0.36	50	$\hat{F}_{NP}$	0.018	0.086	0.081	0.008	0.879	0.053	0.879
				$\hat{F}_{emp}$	0.040	0.054	0.051	0.005	0.471	0.005	0.427
				$\hat{F}_{SP}$	0.008	0.038	0.032	0.002	0.892	-0.048	0.331
B.3	8	0.22,0.14,0.36	250	$\hat{F}_{NP}$	0.009	0.038	0.040	0.002	0.928	-0.051	0.328
				$\hat{F}_{emp}$	0.042	0.024	0.023	0.003	0.333	0.009	0.189
				$\hat{F}_{SP}$	0.006	0.085	0.068	0.008	0.848	-0.039	0.790
B.4	8	0.33,0.08,0.41	50	$\hat{F}_{NP}$	0.023	0.087	0.085	0.008	0.901	-0.080	0.828
				$\hat{F}_{emp}$	0.077	0.053	0.050	0.011	0.460	0.705	0.453
				$\hat{F}_{SP}$	0.010	0.038	0.032	0.002	0.861	-0.056	0.323
B.4	8	0.33,0.08,0.41	250	$\hat{F}_{NP}$	0.009	0.040	0.040	0.002	0.936	-0.058	0.347
				$\hat{F}_{emp}$	0.075	0.024	0.023	0.008	0.304	0.701	0.200

Web Table 3

Table 3: Simulation results under violation of the independence assumption

Survival and truncation times simulated from a normal copula with correlations ρ_{LT}*,* ρ_{LR}*, and* ρ_{TR}*, where* ρ_{XY} *denotes the correlation between random variables* X *and* Y*. The marginal distributions for the survival, left, and right truncation times are set to* $gamma(10,1)$*,* $gamma(4,1.5)$*, and* $gamma(8,1.5)$ *distributions, respectively. The average observed sample size is* $n=250$ *for all models.*

Model	$\rho_{LT},\rho_{LR},\rho_{RT}$	q_L,q_R,q	Estimator	Bias($\hat{F}$)	SD($\hat{F}$)	$\widehat{\text{SE}}(\hat{F})$	MSE($\hat{F}$)	Cov($\hat{F}$)	Bias($\hat{t}_{0.5}$)	SD($\hat{t}_{0.5}$)
			$\hat{F}_{SP}$	0.031	0.042	0.039	0.003	0.865	0.373	0.422
C.1	0.5,0.1,0.1	0.00,0.53,0.0.53	$\hat{F}_{NP}$	0.027	0.043	0.040	0.003	0.883	0.347	0.454
			$\hat{F}_{emp}$	0.142	0.025	0.019	0.023	0.331	-1.294	0.202
			$\hat{F}_{SP}$	0.028	0.038	0.038	0.003	0.847	-0.236	0.311
C.2	-0.5,0.1,-0.1	0.04,0.52,0.56	$\hat{F}_{NP}$	0.028	0.039	0.037	0.003	0.832	-0.211	0.319
			$\hat{F}_{emp}$	0.146	0.024	0.019	0.025	0.331	-1.295	0.182
			$\hat{F}_{SP}$	0.173	0.028	0.034	0.034	0.048	-1.436	0.199
C.3	-0.1,0.1,-0.5	0.02,0.52,0.54	$\hat{F}_{NP}$	0.174	0.028	0.027	0.035	0.021	-1.437	0.204
			$\hat{F}_{emp}$	0.218	0.021	0.016	0.054	0.235	-1.786	0.163
			$\hat{F}_{SP}$	0.157	0.027	0.035	0.030	0.134	-1.301	0.185
C.4	-0.5,0.1,-0.5	0.04,0.52,0.56	$\hat{F}_{NP}$	0.157	0.027	0.027	0.029	0.089	-1.289	0.190
			$\hat{F}_{emp}$	0.204	0.020	0.017	0.050	0.246	-1.663	0.147

Web Table 4

Table 4: Simulation results when $R = L + d_0$ under violation of the independence assumption.

Survival and left truncation times simulated from a normal copula with correlation ρ_{LT}. The marginal distributions for the survival times and left truncation times are set to gamma$(10, 1)$ and gamma$(4.5, 1.5)$, respectively. In all models, we simulate $R = L + d_0$, where d_0 is constant. The average observed sample size is $n = 250$ for all models.

Model	d_0	ρ_{LT}	q_L, q_R, q	Estimator	Bias($\hat{F}$)	SD($\hat{F}$)	$\widehat{\text{SE}}(\hat{F})$	MSE($\hat{F}$)	Cov($\hat{F}$)	Bias($\hat{t}_{0.5}$)	SD($\hat{t}_{0.5}$)
				$\hat{F}_{SP}$	0.060	0.035	0.029	0.006	0.442	-0.168	0.234
D.1	6	-0.5	0.26,0.31,0.57	$\hat{F}_{NP}$	0.061	0.036	0.035	0.006	0.525	-0.160	0.242
				$\hat{F}_{emp}$	0.098	0.020	0.020	0.013	0.245	-0.326	0.139
				$\hat{F}_{SP}$	0.041	0.060	0.038	0.006	0.724	0.416	0.791
D.2	6	0.5	0.14,0.18,0.33	$\hat{F}_{NP}$	0.035	0.060	0.059	0.005	0.933	0.346	0.724
				$\hat{F}_{emp}$	0.058	0.026	0.024	0.005	0.353	-0.530	0.215
				$\hat{F}_{SP}$	0.055	0.030	0.027	0.004	0.444	0.142	0.214
D.3	8	-0.5	0.26,0.19,0.45	$\hat{F}_{NP}$	0.054	0.031	0.030	0.004	0.487	0.154	0.224
				$\hat{F}_{emp}$	0.081	0.021	0.021	0.008	0.327	0.164	0.157
				$\hat{F}_{SP}$	0.047	0.042	0.034	0.005	0.639	-0.432	0.422
D.4	8	0.5	0.14,0.06,0.21	$\hat{F}_{NP}$	0.048	0.043	0.044	0.005	0.739	-0.449	0.428
				$\hat{F}_{emp}$	0.017	0.026	0.025	0.001	0.336	-0.092	0.224