

Web-based Supporting Materials for “A Bayesian Model for Misclassified Binary Outcomes and Correlated Survival Data with Applications to Breast Cancer” by Sheng Luo, Min Yi, Xuelin Huang, and Kelly K. Hunt

Table 1: The estimates of the posterior mean (PM), standard error (SE), standard deviation (SD), and coverage probabilities (CP) of 95% credible intervals based on logistic regression (LR), the reduced model, and the proposed model when no misclassification is present and conditional independence is assumed.

	True pars	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$p_1$	$q_1$	$p_2$	$q_2$
LR	PM	0.305	0.007	-1.797	0.605	—	—	—	—
	SD	0.131	0.009	0.270	0.381	—	—	—	—
	SE	0.127	0.009	0.265	0.380	—	—	—	—
	CP	0.940	0.950	0.990	0.930	—	—	—	—
Reduced model	PM	0.307	0.007	-1.822	0.618	0.995	0.995	0.995	0.995
	SD	0.132	0.009	0.276	0.390	0.005	0.005	0.005	0.005
	SE	0.125	0.009	0.266	0.358	0.005	0.005	0.005	0.005
	CP	0.940	0.940	0.970	0.910	—	—	—	—
Proposed model	PM	0.308	0.007	-1.824	0.618	0.995	0.995	0.995	0.995
	SD	0.132	0.009	0.276	0.390	0.005	0.005	0.005	0.005
	SE	0.126	0.009	0.265	0.359	0.005	0.005	0.005	0.005
	CP	0.950	0.950	0.960	0.900	—	—	—	—

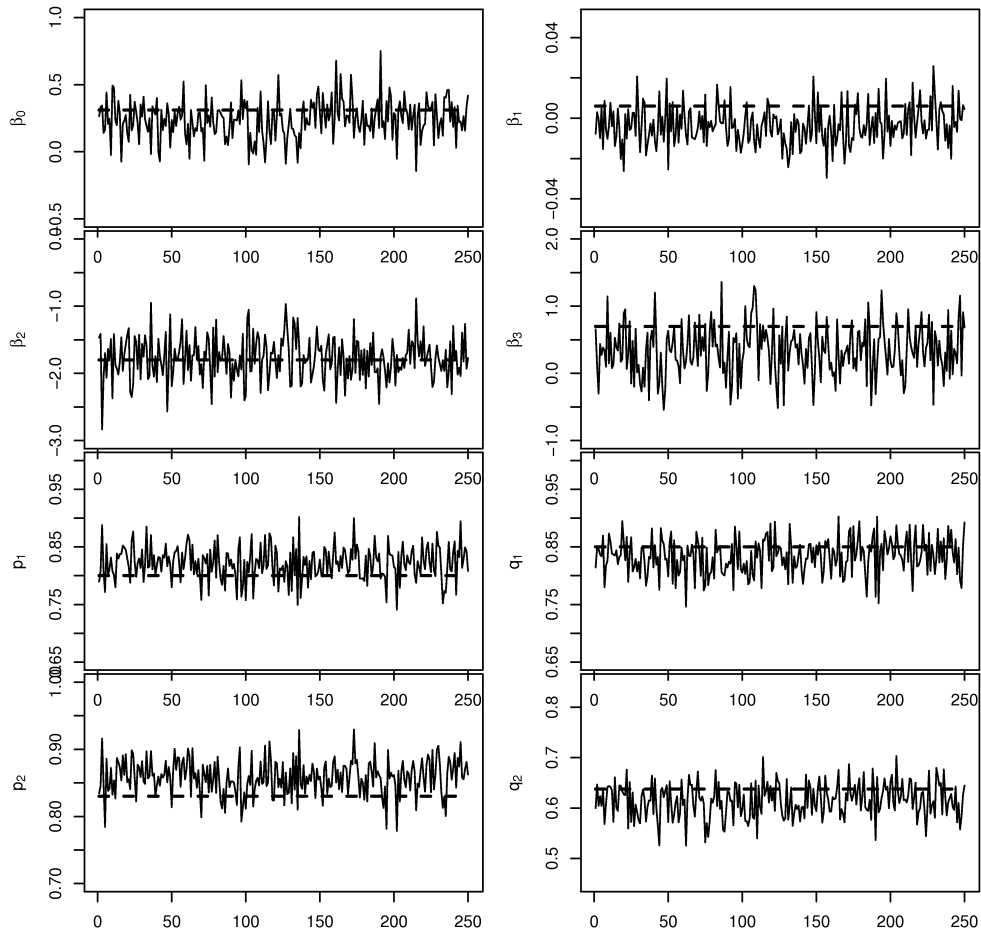


Figure 1: MCMC histories for 8 parameters of interest in one simulated dataset. The horizontal dash lines denote the true parameter values.

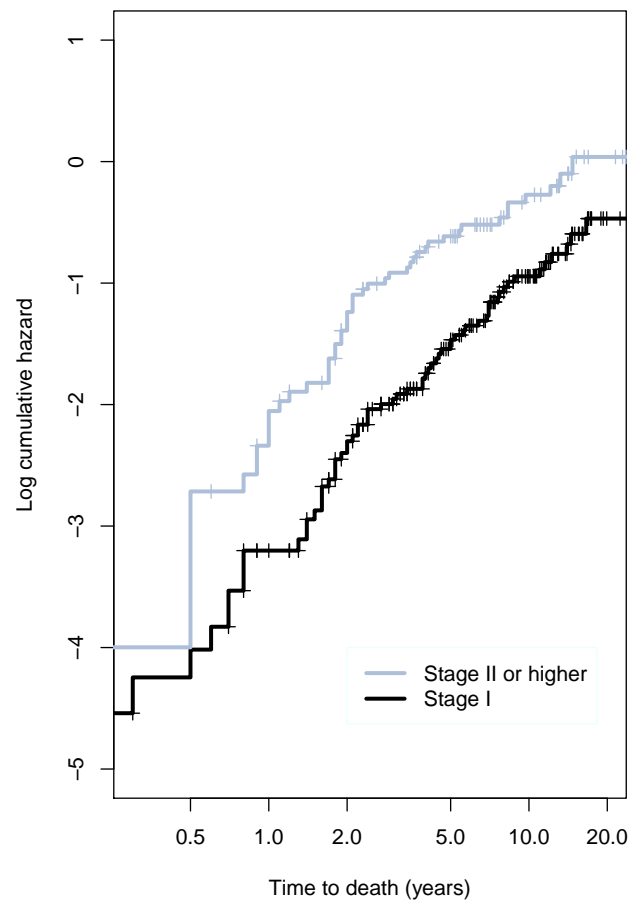
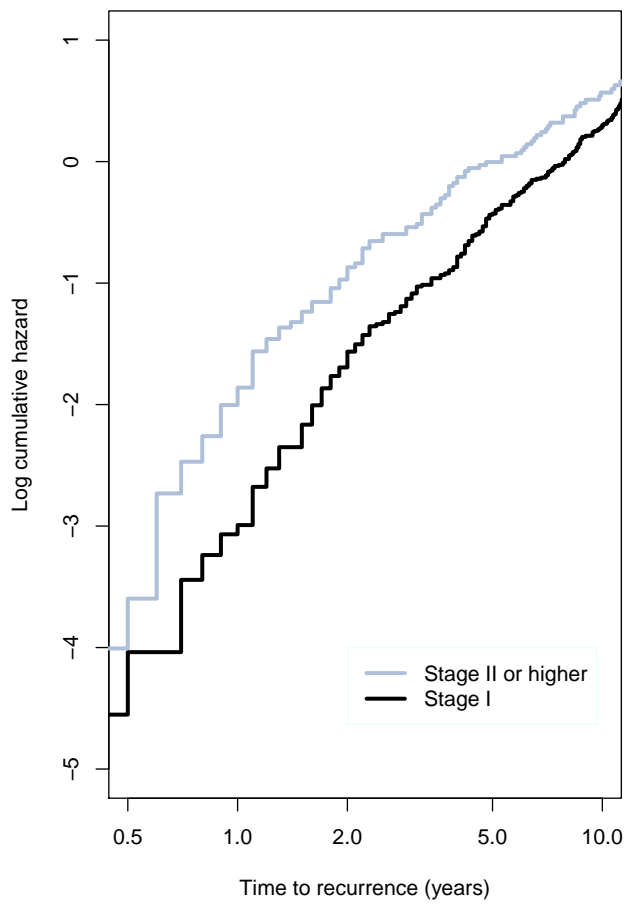


Figure 2: Log cumulative hazard curves for two tumor stage groups.

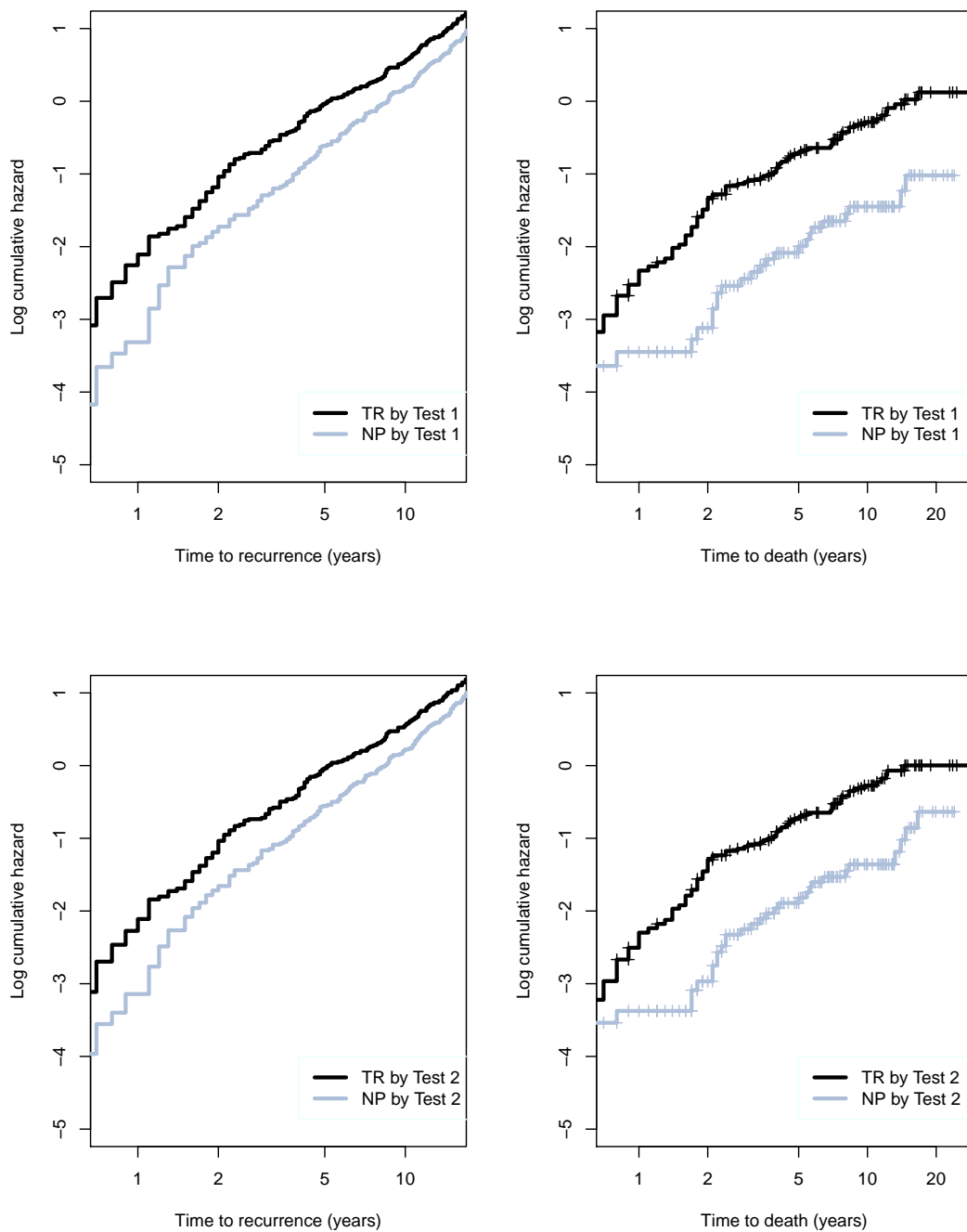


Figure 3: Log cumulative hazard curves of time to relapse (left panels) and time from IBTR to death (right panels) for patients classified as NP or TR by test 1 (top panels) and test 2 (bottom panels).

Table 2: Model comparison statistics for the MDACC dataset. The variables in the column labeled Model are in the binomial regression model with misclassification, while  $x_1$  and  $x_4$  are included in the survival models. DIC: deviance information criterion;  $\overline{D}$ : the posterior mean of the deviance;  $p_D$ : the effective number of parameters;  $x_1$ : age at diagnosis;  $x_2$ : distant recurrence;  $x_3$ : contralateral;  $x_4$ : tumor stage.

Baseline hazard	Model	DIC	$\overline{D}$	$p_D$
Piecewise constant	$x_1, x_2$	3493.9	3398.7	92.2
	$x_1, x_3$	3721.0	3671.0	50.0
	$x_1, x_4$	3740.0	3696.7	43.3
	$x_1, x_2, x_3$	3480.0	3377.9	102.5
	$x_1, x_2, x_4^*$	<b>3458.4</b>	<b>3374.7</b>	<b>83.7</b>
Weibull distribution	$x_1, x_2, x_4$	3500.5	3391.1	109.4

\*Boldface indicates the preferred model.