

Web-based Supplementary Materials for “A Bayesian Semi-parametric Survival Model with Longitudinal Markers”

Song Zhang, Peter Müller and Kim-Anh Do

Web Appendix A. A Pseudo Prior Model

From L_{vi0} with censored observations the two values taken by ω_{vi} lead to two models of different dimensions. If $\omega_{vi} = 0$, we have a model with T_{vi} being a random parameter. In contrast, T_{vi} is fixed at $T_{vi} = t_c$ if $\omega_{vi} = 1$. Such a change in dimension complicates posterior simulation (Green, 1995). We use the pseudo prior approach by Carlin and Chib (1995) to avoid this complication. In other words, we augment the smaller probability model under $\omega_{vi} = 1$ by defining a prior probability model for a hypothetical T_{vi} (but keep t_c in the regression for \mathbf{y}_{vi}). The new variable T_{vi} has no meaningful interpretation under $\omega_{vi} = 1$. It is only introduced to match the model dimensions. The augmented likelihood factor under the new model is

$$L_{vi0}^* = \{[\mathbf{y}_{vi} | t_c, \mathbf{\Psi}] p_v \pi_{vi}(T_{vi})\}^{\omega_{vi}} \cdot \{[\mathbf{y}_{vi} | T_{vi}, \mathbf{\Psi}] (1 - p_v) g_v(T_{vi}) I(T_{vi} > t_{vi})\}^{1-\omega_{vi}}. \quad (1)$$

Here $\pi_{vi}(T_{vi})$ is a pseudo prior for T_{vi} when $\omega_{vi} = 1$. It is a conveniently chosen linking density such that the two models implied by $\omega_{vi} = 0/1$ have the same dimension. The equivalence between L_{vi0}^* and L_{vi0} is obvious when $\omega_{vi} = 0$. When $\omega_{vi} = 1$, the equivalence can be verified by integrating (1) with regard to T_{vi} . As the name “pseudo prior” suggests, $\pi_{vi}(T_{vi})$ has no

effect on model inference. However, a poor choices of the linking density may lead to poor mixing of the posterior Markov chain Monte Carlo simulation. In our implementation we follow the recommendation of Carlin and Chib (1995) and base the specification of pseudo priors $\pi_{vi}(T_{vi})$ on a preliminary data analysis. First we fit a model without cure where $\omega_{vi} = 0$ for all subjects. Under this model there is no dimensional change. Then $\pi_{vi}(T_{vi})$ is specified to mimic the marginal posterior density of T_{vi} under the simplified model. Specifically, we assume $\pi_{vi}(T_{vi})$ to be normal and match the first two moments. Finally, under the pseudo priors approach the marginal posterior distribution of T_{vi} is meaningless. Only the posterior conditional density of T_{vi} given $\omega_{vi} = 0$ is of interest.

Web Appendix B. The Polya Tree Prior

For reference, we give a brief review of PT models. More details can be found in Lavine (1992, 1994) and Mauldin et al. (1992). Let $\epsilon = \epsilon_1 \cdots \epsilon_m \in E^m$ denote a binary sequence of length m . For example, $E^1 = \{0, 1\}$ and $E^2 = \{00, 01, 10, 11\}$. The definition of the PT prior requires two parameters, a nested sequence of partitions $\Pi = \{B_0, B_1, B_{00}, B_{01}, \dots, B_{\epsilon_0}, B_{\epsilon_1}, \dots\}$ of the sample space S ,

$$S = B_0 \cup B_1, B_0 = B_{00} \cup B_{01}, B_1 = B_{10} \cup B_{11}, \dots, B_\epsilon = B_{\epsilon_0} \cup B_{\epsilon_1}, \dots,$$

and parameters $\mathcal{A} = \{\alpha_0, \alpha_1, \alpha_{00}, \alpha_{01}, \dots, \alpha_{\epsilon_0}, \alpha_{\epsilon_1}, \dots\}$ that define a sequence of random variables $Y_{\epsilon_0} \sim Be(\alpha_{\epsilon_0}, \alpha_{\epsilon_1})$ and $Y_{\epsilon_1} = 1 - Y_{\epsilon_0}$, independently across ϵ . We say that a random probability measure G has a PT prior, $G \sim PT(\Pi, \mathcal{A})$, if the random probability $G(B_{\epsilon_0} | B_\epsilon)$ is defined by $G(B_{\epsilon_0} | B_\epsilon) \equiv Y_{\epsilon_0}$. This implies $G(B_{\epsilon_1, \dots, \epsilon_m}) = \prod_{j=1}^m Y_{\epsilon_1, \dots, \epsilon_j}$. We can center G around a given distribution \tilde{G} , i.e., $E(G(B)) = \tilde{G}(B)$, by setting $\alpha_{\epsilon_0} = \alpha_{\epsilon_1}$ and taking the partition Π at level m to coincide with quantiles $\tilde{G}^{-1}(k/2^m)$, $k = 0, 1, \dots, 2^m$. That is, for any $\epsilon \in E^m$,

$$B_\epsilon = (\tilde{G}^{-1}(k/2^m), \tilde{G}^{-1}((k+1)/2^m)) \tag{2}$$

for some k in $\{0, 1, \dots, 2^m - 1\}$.

The family \mathcal{A} determines how much G varies around \tilde{G} . It has a similar role as the precision parameter in a Dirichlet process prior. Berger and Guglielmi (2001) considered a family of the form $\alpha_{\epsilon_1, \dots, \epsilon_m} = c \cdot \rho(m)$, where $\rho(m) = m^2, m^3, 2^m, 4^m$ or 8^m , and $c > 0$ is a constant. In general, any $\rho(m)$ such that $\sum_{m=1}^{\infty} \rho(m)^{-1} < \infty$ guarantees G to be absolutely continuous. For example, $\rho(m) = m^{1+\eta}$ or $\rho(m) = (1 + \eta)^m$ for $\eta > 0$ satisfies the above condition.

A technically convenient property of PT priors is the conjugacy under random sampling. Let $n_\epsilon(\mathbf{T})$ be the number of elements of \mathbf{T} contained in B_ϵ . The posterior distribution of G given \mathbf{T} is again a PT, $G | \mathbf{T} \sim PT(\Pi, \mathcal{A}')$, where $\mathcal{A}' = \{\alpha'_\epsilon\}$ with $\alpha'_\epsilon = \alpha_\epsilon + n_\epsilon(\mathbf{T})$.

Another useful property is the following closed form expression for the predictive density function of $(T_n | T_1, \dots, T_{n-1})$, marginalized with respect to G . Let $\mathbf{T}_{(-i)} = \{T_j : j \neq i\}$, let $\epsilon(j, T_i)$ denote the index $\epsilon_1 \dots \epsilon_j \in E^j$ such that $T_i \in B_{\epsilon_1 \dots \epsilon_j}$, and let $\tilde{g}(\cdot)$ be the density function of \tilde{G} . Assume that the partition Π is specified as in (2), and \mathcal{A} is specified such that for every $\epsilon \in E^m$, $\alpha_\epsilon = c \cdot m^2$. Define M_i to be the smallest integer such that $n_{\epsilon(M_i, T_i)}(\mathbf{T}_{(-i)}) = 0$. The marginal predictive distribution can be computed exactly:

$$[T_n | T_1, \dots, T_{n-1}] = \left\{ \prod_{j=1}^{M_n} \frac{c j^2 + n_{\epsilon(j, T_n)}(\mathbf{T}_{(-n)})}{2c j^2 + n_{\epsilon(j-1, T_n)}(\mathbf{T}_{(-n)})} \right\} 2^{M_n} \tilde{g}(T_n). \quad (3)$$

See, for example, Hanson and Johnson (2002).

The conditional cumulative probability marginalized with respect to G , $[T_n < t | T_1, \dots, T_{n-1}]$ can also be evaluated exactly. We introduce the following notation. Let M_i^* be the smallest integer such that $n_{\epsilon(M_i^*, t)}(\mathbf{T}_{(-i)}) = 0$. Let $\mathcal{D}_i(t) = \{\epsilon : \epsilon \in E^{M_i^*} \text{ and } B_\epsilon < t\}$ be the set of indices for partitions defined at level M_i^* and on the left of t . Here $B_\epsilon < t$ indicates that the upper bound of B_ϵ is smaller than t . For a partition B_ϵ , $\epsilon \in \mathcal{D}_i(t)$, we define $\epsilon^*(j, B_\epsilon)$, $j = 1, \dots, M_i^*$, to be the sequence of indices such that $B_{\epsilon^*(1, B_\epsilon)} \supseteq B_{\epsilon^*(2, B_\epsilon)} \supseteq \dots \supseteq B_{\epsilon^*(M_i^*, B_\epsilon)} = B_\epsilon$. We further define $B_i(t) = B_{\epsilon(M_i^*, t)} \cap (-\infty, t)$. Then the marginal cumulative

distribution function is

$$\begin{aligned}
[T_n < t \mid T_1, \dots, T_{n-1}] &= \sum_{\epsilon \in \mathcal{D}_n(t)} E(G(B_\epsilon)) + E(G(B_i(t))) \\
&= \sum_{\epsilon \in \mathcal{D}_n(t)} \prod_{j=1}^{M_n^*} \frac{cj^2 + n_{\epsilon^*(j, B_\epsilon)}(\mathbf{T}_{(-n)})}{2cj^2 + n_{\epsilon^*(j-1, B_\epsilon)}(\mathbf{T}_{(-n)})} \\
&\quad + \prod_{j=1}^{M_n^*} \frac{cj^2 + n_{\epsilon(j, t)}(\mathbf{T}_{(-n)})}{2cj^2 + n_{\epsilon(j-1, t)}(\mathbf{T}_{(-n)})} 2^{M_i^*} \tilde{G}(B_i(t)). \tag{4}
\end{aligned}$$

The term $\sum_{\epsilon \in \mathcal{D}_n(t)} E(G(B_\epsilon))$ can be computed more efficiently when combining B_ϵ , $\epsilon \in \mathcal{D}_i(t)$, into partitions defined at higher levels of the Pólya tree. The marginalized conditional survival probability, $[T_n > t \mid T_1, \dots, T_{n-1}]$, can be computed in the same fashion. Expression (4) is useful in the computation of CPO.

Web Appendix C. Posterior Sampling Scheme

Posterior MCMC simulation is built on sampling from the following conditional posterior distributions and other transition probabilities.

The simulation of the full conditional posterior distribution of Ψ depends on the prior model assumed. Different sampling strategies have been discussed by Gelman et al. (2003).

Under the pseudo priors setup, simulation of ω_v^0 is straightforward. The full conditional posterior distribution of an unknown ω_{vi} is a Bernoulli(p_{vi}^*) with

$$p_{vi}^* = \frac{[\mathbf{y}_{vi} \mid t_c, \Psi] p_v \pi_{vi}(T_{vi})}{[\mathbf{y}_{vi} \mid t_c, \Psi] p_v \pi_{vi}(T_{vi}) + [\mathbf{y}_{vi} \mid T_{vi}, \Psi] (1 - p_v) [T_{vi} \mid \mathbf{T}_{v(-i)}^s]}.$$

Here $\mathbf{T}_{v(-i)}^s$ denotes the set of observed and unobserved TTP in the susceptible group except T_{vi} . For censored subjects ($d_{vi} = 0$), T_{vi} needs to be simulated. Given that $\omega_{vi} = 1$, we simulate T_{vi} from the pseudo prior, i.e., $T_{vi} \sim \pi_{vi}(T_{vi})$. Given that $\omega_{vi} = 0$, we simulate T_{vi} by Acceptance-Rejection sampling (Robert and Casella, 2003). The full conditional distribution of T_{vi} is proportional to $[\mathbf{y}_{vi} \mid T_{vi}, \Psi] I(T_{vi} > t_{vi}) [T_{vi} \mid \mathbf{T}_{v(-i)}^s]$. Because $[\mathbf{y}_{vi} \mid T_{vi}, \Psi]$ is bounded

and easy to evaluate, we can propose values based on $[T_{vi} | \mathbf{T}_{v(-i)}^s]I(T_{vi} > t_{vi})$, and decide whether to accept or reject the proposal based on $[\mathbf{y}_{vi} | T_{vi}, \mathbf{\Psi}]$. The proposed values are generated from $[T_{vi} | \mathbf{T}_{v(-i)}^s]$ under the constraint that $T_{vi} > t_{vi}$. The definition of $[T_{vi} | \mathbf{T}_{v(-i)}^s]$ is given in (3), which is the posterior predictive distribution under a PT prior. Lavine (1992) describes how to generate random samples from the posterior predictive distribution.

The full conditional distribution of p_v is proportional to

$$p_v^{a_p + \sum_{i=1}^{n_v} \omega_{vi} - 1} (1 - p_v)^{b_p + \sum_{i=1}^{n_v} (1 - \omega_{vi}) - 1},$$

which is again a Beta distribution.

Web Appendix D. The Computation of CPO

The computation of CPO differs for censored and uncensored subjects. We assume that $d_{vi} = 0$ for $i = 1, \dots, n_{v0}$, and $d_{vi} = 1$ for $i = n_{v0} + 1, \dots, n_v$. First we derive CPO for uncensored cases, i.e., $d_{vi} = 1$. Defining $\mathbf{t}_{v(-i)}^1 = \mathbf{t}_v^1 / t_{vi}$ and integrating with respect to G_v , we have

$$\begin{aligned} CPO_{vi} &= \int [\mathbf{y}_{vi} | t_{vi}, \mathbf{\Psi}] [t_{vi} | \mathbf{t}_{v(-i)}^1, \mathbf{T}_v^0] (1 - p_v) [\mathbf{\Psi}, \mathbf{T}_v^0, p_v | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}] d\mathbf{\Psi} d\mathbf{T}_v^0 dp_v \\ &= \int [\mathbf{y}_{vi} | t_{vi}, \mathbf{\Psi}] [t_{vi} | \mathbf{t}_{v(-i)}^1, \mathbf{T}_v^0] (1 - p_v) [\mathbf{\Lambda} | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}] d\mathbf{\Lambda}. \end{aligned} \quad (5)$$

Note that when marginalized with regard to G_v , the likelihood contribution from subject (v, i) is $[\mathbf{y}_{vi} | t_{vi}, \mathbf{\Psi}] [t_{vi} | \mathbf{t}_{v(-i)}^1, \mathbf{T}_v^0] (1 - p_v)$. We have the second equation because $[\mathbf{\Psi}, \mathbf{T}_v^0, p_v | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}]$ is the marginal distribution of $(\mathbf{\Psi}, \mathbf{T}_v^0, p_v)$ obtained from $[\mathbf{\Lambda} | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}]$. Using the fact that

$$[\mathbf{\Lambda} | \mathbf{Y}, \mathbf{t}, \mathbf{d}] \propto [\mathbf{y}_{vi} | t_{vi}, \mathbf{\Psi}] [t_{vi} | \mathbf{t}_{v(-i)}^1, \mathbf{T}_v^0] (1 - p_v) [\mathbf{\Lambda} | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}],$$

we can evaluate (5) through an importance sampling scheme. The full posterior distribution, $[\mathbf{\Lambda} | \mathbf{Y}, \mathbf{t}, \mathbf{d}]$, serves as the importance sampling density, and the reciprocal of the likelihood

contribution from (v, i) serves as the importance sampling weight. Specifically, we estimate CPO_{vi} by

$$CPO_{vi} \approx \left\{ \frac{1}{K} \sum_{k=1}^K \frac{1}{[\mathbf{y}_{vi} | t_{vi}, \Psi^{(k)}][t_{vi} | \mathbf{t}_{v(-i)}^1, \mathbf{T}_v^{0(k)}](1 - p_v^{(k)})} \right\}^{-1},$$

where $(\Psi^{(k)}, \mathbf{T}_v^{0(k)}, p_v^{(k)})$ is the k th sample from the full posterior distribution, $[\Lambda | \mathbf{Y}, \mathbf{t}, \mathbf{d}]$, given all observations.

For censored cases, i.e., $d_{vi} = 0$, the computation of CPO is more complicated. Integrating over G_v , the augmented likelihood factor is

$$[\mathbf{y}_{vi} | T_{vi}, \Psi] I(T_{vi} > t_{vi}) [T_{vi} | \omega_{vi}, \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0] [\omega_{vi} | p_v],$$

where $\mathbf{T}_{v(-i)}^0 = \mathbf{T}_v^0 / T_{vi}$ and $[T_{vi} | \omega_{vi}, \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0] = \{\delta(t_c)\}^{\omega_{vi}} [T_{vi} | \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0]^{1-\omega_{vi}}$. Thus

$$CPO_{vi} = \int [\mathbf{y}_{vi} | T_{vi}, \Psi] I(T_{vi} > t_{vi}) [T_{vi} | \omega_{vi}, \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0] [\omega_{vi} | p_v] \\ \cdot [\Psi, \mathbf{T}_{v(-i)}^0, p_v | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}] d\Psi d\mathbf{T}_v^0 d\omega_{vi} dp_v.$$

Note that $[\Psi, \mathbf{T}_{v(-i)}^0, p_v | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}]$ is obtained from $[\Lambda_{(-vi)} | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}]$ by marginalizing over parameters other than $(\Psi, \mathbf{T}_{v(-i)}^0, p_v)$. Here $\Lambda_{(-vi)}$ is the set of model parameters based on $(\mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)})$, and $[\Lambda_{(-vi)} | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}]$ is the full posterior distribution as defined by Expression (2) in Zhang et al. (2008). Define

$$P(T > t_{vi} | \mathbf{T}_{v(-i)}^0, p_v) = \int I(T_{vi} > t_{vi}) [T_{vi} | \omega_{vi}, \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0] [\omega_{vi} | p_v] d\omega_{vi} dT_{vi} \\ = \int I(T_{vi} > t_{vi}) \{p_v \delta(t_c) + (1 - p_v) [T_{vi} | \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0]\} dT_{vi} \\ = p_v + (1 - p_v) [T > t_{vi} | \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0],$$

where $[T > t_{vi} | \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0]$ is the conditional survival probability of T_{vi} marginalized with respect to G_v , as is discussed in (4). We then define $f(\Lambda)$ as the product of two density functions

$$f(\Lambda) = \left\{ \frac{I(T_{vi} > t_{vi}) [T_{vi} | \omega_{vi}, \mathbf{t}_v^1, \mathbf{T}_{v(-i)}^0] [\omega_{vi} | p_v]}{P(T > t_{vi} | \mathbf{T}_{v(-i)}^0, p_v)} \right\} \cdot [\Lambda_{(-vi)} | \mathbf{Y}_{(-vi)}, \mathbf{t}_{(-vi)}, \mathbf{d}_{(-vi)}].$$

Thus we have

$$CPO_{vi} = \int [\mathbf{y}_{vi} | T_{vi}, \Psi] P(T > t_{vi} | \mathbf{T}_{v(-i)}^0, p_v) f(\Lambda) d\Lambda.$$

Since $[\Lambda | \mathbf{Y}, \mathbf{t}, \mathbf{d}] \propto [\mathbf{y}_{vi} | T_{vi}, \Psi] P(T > t_{vi} | \mathbf{T}_{v(-i)}^0, p_v) f(\Lambda)$, an importance sampling scheme can be employed to evaluate CPO, with $[\Lambda | \mathbf{Y}, \mathbf{t}, \mathbf{d}]$ being the importance sampling density and $\{[\mathbf{y}_{vi} | T_{vi}, \Psi] P(T > t_{vi} | \mathbf{T}_{v(-i)}^0, p_v)\}^{-1}$ being the importance sampling weight, i.e.,

$$CPO_{vi} \approx \left\{ \frac{1}{K} \sum_{k=1}^K \frac{1}{[\mathbf{y}_{vi} | T_{vi}^{(k)}, \Psi^{(k)}] P(T > t_{vi} | \mathbf{T}_{v(-i)}^{0(k)}, p_v^{(k)})} \right\}^{-1}. \quad (6)$$

Here $(\Psi^{(k)}, T_{vi}^{(k)}, \mathbf{T}_{v(-i)}^{0(k)}, p_v^{(k)})$ are the k th sample from the full posterior distribution $[\Lambda | \mathbf{Y}, \mathbf{t}, \mathbf{d}]$.

Web Appendix E. Plots

- Figure 1 plots the observed and fitted PSA profiles for four randomly selected patients.
- Figure 2 validates the survival and cure aspects of the proposed model based on subject specific martingale residuals (Barlow and Prentice, 1988; Therneau et al., 1990; Lin et al., 2002), which is defined by $e_{vi} = d_{vi} - H_{vi}$. Here H_{vi} is the individual cumulative hazard up to t_{vi} . The residuals can be interpreted as the difference over $[0, t_{vi}]$ in the observed number of events and the expected number given the model. In general, the residuals are scattered horizontally over age (with three outliers), suggest that the proposed model is sufficient.
- Figure 3 shows the posterior variability of G_v by plotting ten random samples from its posterior distribution.
- Figure 4 shows the intial drop and duration induced by the AA/CH treatments. A larger value of η_v suggests a deeper initial drop in PSA level. On the other hand, the

larger the value of ϕ_{1v} , the sooner the treatment effect wears out. We plot $l_v(t) = \eta_v[\exp(-\phi_{1v}t) - 1]$ in Figure 4.

References

- Barlow, W. E. and Prentice, R. L. (1988), ‘Residuals for relative risk regression’, *Biometrika* **75**, 65–74.
- Berger, J. O. and Guglielmi, A. (2001), ‘Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives’, *Journal of the American Statistical Association* **96**(453), 174–184.
- Carlin, B. P. and Chib, S. (1995), ‘Bayesian model choice via Markov chain Monte Carlo methods’, *Journal of the Royal Statistical Society, Series B: Methodological* **57**, 473–484.
- Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2003), *Bayesian Data Analysis*, Chapman & Hall.
- Green, P. J. (1995), ‘Reversible jump Markov chain Monte Carlo computation and Bayesian model determination’, *Biometrika* **82**, 711–732.
- Hanson, T. and Johnson, W. O. (2002), ‘Modeling regression error with a mixture of Polya trees’, *Journal of the American Statistical Association* **97**(460), 1020–1033.
- Lavine, M. (1992), ‘Some aspects of Polya tree distributions for statistical modelling’, *The Annals of Statistics* **20**, 1222–1235.
- Lavine, M. (1994), ‘More aspects of Polya tree distributions for statistical modelling’, *The Annals of Statistics* **22**, 1161–1176.
- Lin, H., Turnbull, B. W., McCulloch, C. E. and Slate, E. H. (2002), ‘Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal

prostate-specific antigen readings and prostate cancer’, *Journal of the American Statistical Association* **97**(457), 53–65.

Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992), ‘Polya trees and random distributions’, *The Annals of Statistics* **20**, 1203–1221.

Robert, C. P. and Casella, G. (2003), *Monte Carlo Statistical Methods*, Springer.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990), ‘Martingale-based residuals for survival models’, *Biometrika* **77**, 147–160.

Zhang, S., Müller, P. and Do, K.-A. (2008), ‘A Bayesian semi-parametric survival model with longitudinal markers’, *submitted to Biometrics* .

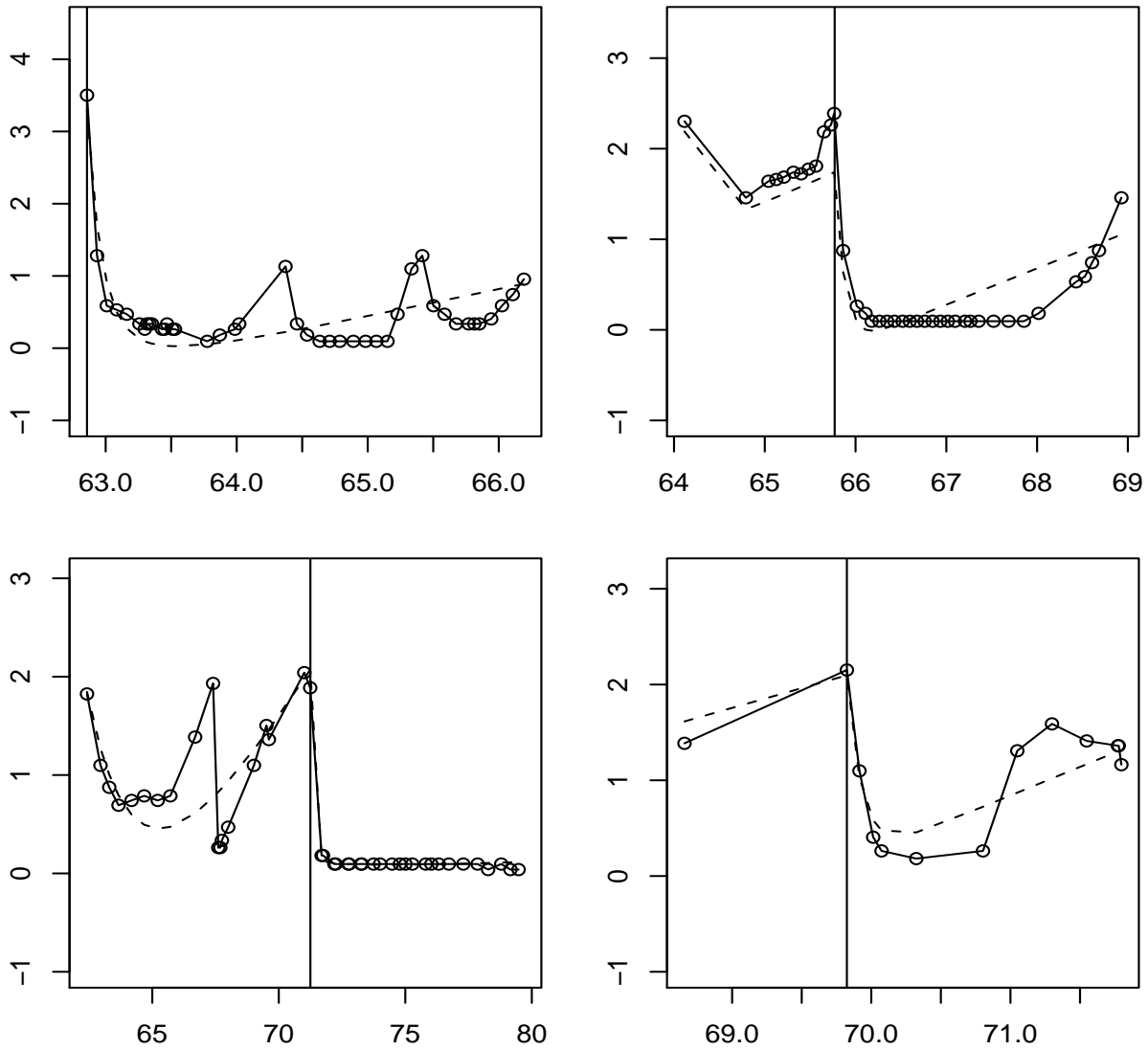


Figure 1: The observed longitudinal profiles and fitted values of 4 randomly selected patients. The vertical axis indicates $\log(\text{PSA} + 1)$, and the horizontal axis is age in years. Each point denotes a PSA measurement. The dotted lines plot fitted values of the longitudinal profiles. The vertical line marks the initiation of the AA/CH therapy.

Residual for the Survival Model

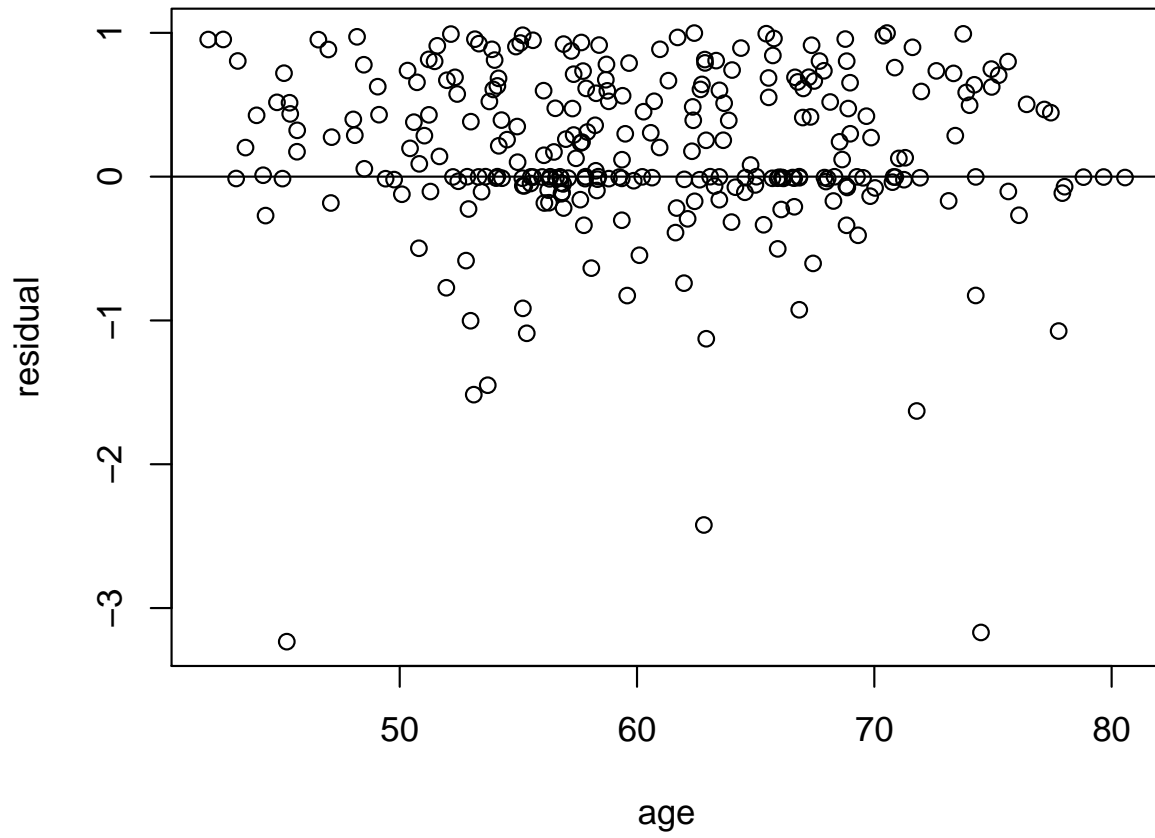


Figure 2: The martingale residual for the survival model versus age. In general, the residuals are scattered horizontally over age.

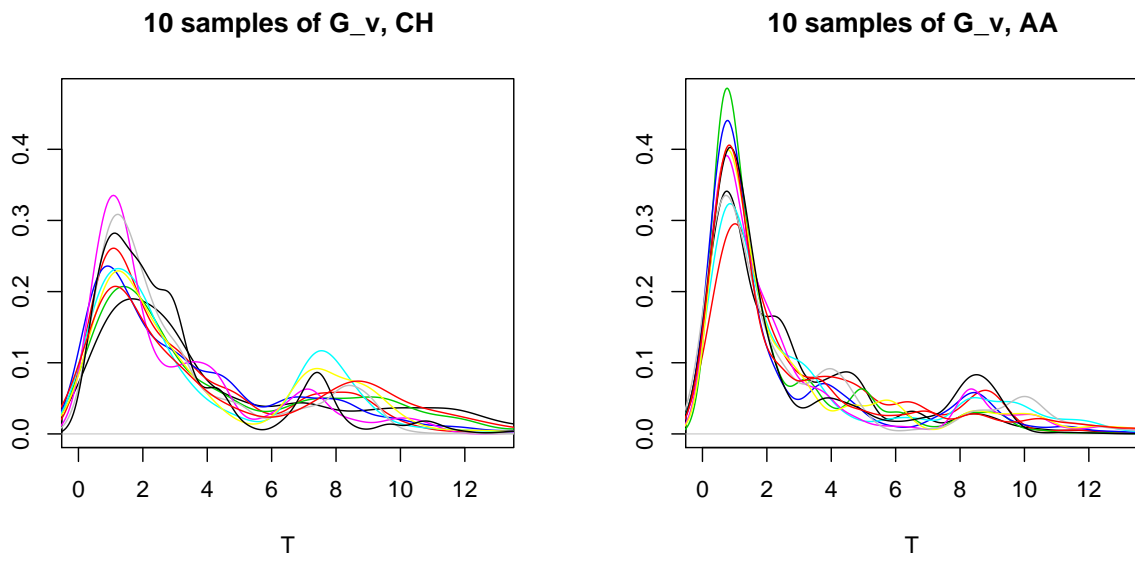


Figure 3: Ten random samples from $[G_v | \mathbf{Y}, \mathbf{t}, \mathbf{d}]$. The horizontal axis indicates years after treatment.

Initial Drop and Duration

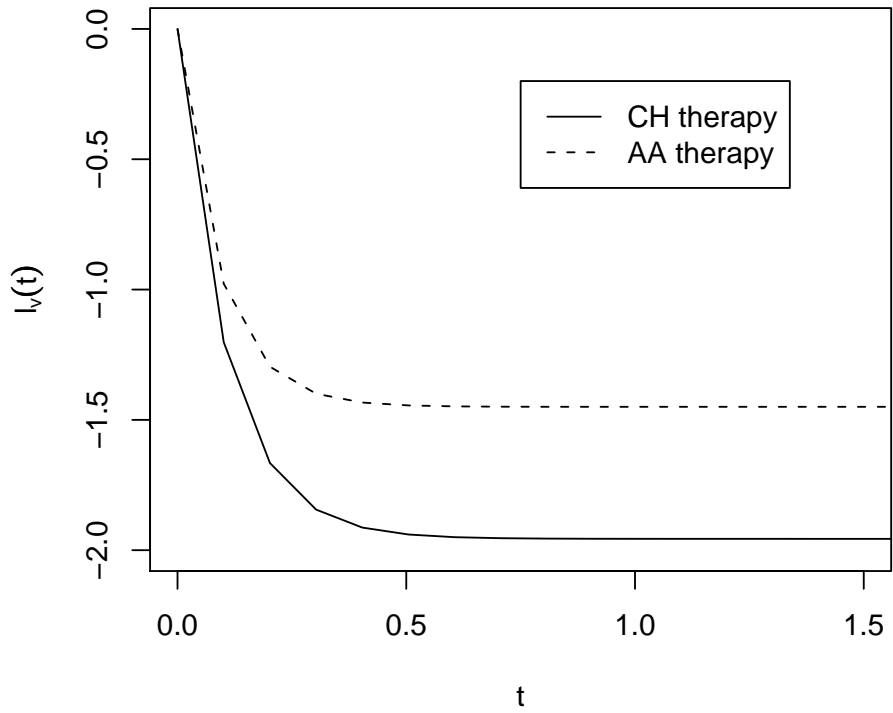


Figure 4: The initial slope and duration of treatment effect modeled by η_v and ϕ_{1v} . Here t is the time in years from the start of treatment v and $l_v(t) = \eta_v\{\{\exp(-\phi_{1v}t) - 1\}\}$.