**Web-based Supplementary Materials for "Bayesian nonparametric estimation of targeted agent effects on biomarker change to predict clinical outcome," by Rebecca Graziani, Michele Guindani, and Peter F. Thall.**

## Prostate Cancer Trial Data

*Testing Assumptions.* We test the validity of the simplifying assumption of no association between the distributions of the PDGFR values $X$ and $Y$ and the other covariates $Z$ in the PFS model, by regressing the individual pre- and post- treatment mean values on hemoglobin levels and increase in prostate antigen levels. No significant association was revealed by such analysis. Here, we report the results of the marginal regression of the mean values on each variable. Multivariable regressions confirmed the results of the marginal analyses.

[Table 1 about here.]

*Time-varying effect of PDGFR inhibition.* Although PDFGR inhibition seems to be associated with increased PFS time in the first few months after therapy, the opposite seems true after some time. An extended Cox proportional hazard model confirms this time-varying effect (Table 2).

[Table 2 about here.]

*Posterior distribution of the biological profile* $\Delta$. Our results suggest that the biological assumption of the trial might have been fallacious, and provide an understanding of the reasons of the negative therapeutic results.

Table 3 expands on Figure 3 in the text and reports the posterior probability that a patient belongs to any of the three clusters for each of the two arms of the study. In particular, the probability of a patient experiencing a shift toward the left of the biomarker profiles after

treatment, corresponding to $E(\Delta|\text{data}) < 0.5$, is slightly higher in the DI arm than in the DP arm; however, the two distributions do not appear to be statistically significantly different according to a $\chi^2$ test of homogeneity (p value=0.2694).

[Table 3 about here.]

## MCMC

In the following we detail the Markov Chain Monte Carlo algorithm used to conduct approximate posterior inference. Following Rodriguez et al. (2010), we use a truncation approximation to the stick-breaking representation of the Dirichlet processes and then resort to methods for computation in finite-mixture models (see Ishwaran and Zarepour 2002; Ishwaran and James 2001). We assume that individuals can be clustered in to $K$ groups and that for each individual the observations on the biomarker level can be clustered into $L$ groups. Two sets of auxiliary variables are introduced: $(\xi_{\mathbf{x}_i}, \xi_{\mathbf{y}_i}),\ i = 1, \ldots, N$ and $(\psi_{X_{ij}}, \psi_{Y_{ik}})\ j = 1, \ldots, n_i,\ , k = 1, \ldots, m_i$, where $\xi_{\mathbf{x}_i} = k$ and $\psi_{X_{ij}} = l$ if $G_{\mathbf{x}_i} = G_k^*$ and $\theta_{X_{ij}} = \theta_{lk}^*$ and $\xi_{\mathbf{y}_i} = k$ and $\psi_{y_{ij}} = l$ if $G_{\mathbf{y}_i} = G_k^*$ and $\theta_{y_{ij}} = \theta_{lk}^*$, to indicate membership to the distributional and observational clusters before and after treatment, respectively. The computation proceeds through the following steps:

(1) Sample $\xi_{\mathbf{x}_i}$, for $i = 1, \ldots, N$, from a multinomial distribution with probabilities:

$$Pr(\xi_{\mathbf{x}_i} = k| \ldots) = q_k^i \propto \pi_k^* \prod_{j=1}^{n_i} \sum_{l=1}^{L} \omega_{lk}^* \, \phi_N(x_{ij}; \theta_{lk}^*),$$

where $\theta_{lk}^* = (\mu_{lk}^*, \sigma_{lk}^*)^T$.

(2) Sample $\psi_{X_{ij}}$, for $j = 1, \ldots, n_i$, from a multinomial distribution with probabilities:

$$Pr(\psi_{X_{ij}} = l| \ldots) = b_{X_{ij}}^l \propto \omega_{\mathbf{x}_i,l}^* \phi_N(x_{ij}; \theta_{\mathbf{x}_i,l}^*).$$

(3) Similarly, sample $\xi_{\mathbf{y}_i}$ for $i = 1, \ldots, N$, from a multinomial distribution with probabilities:

$$Pr(\xi_{\mathbf{y}_i} = k| \ldots) = q_k^i \propto \pi_k^* \prod_{j=1}^{m_i} \sum_{l=1}^{L} \omega_{lk}^* \, \phi_N(y_{ij}; \theta_{lk}^*),$$

(4) Sample $\psi_{y_{ij}}$, for $j = 1, \ldots, m_i$, from a multinomial distribution with probabilities:

$$Pr(\psi_{y_{ij}} = l | \ldots) = b^l_{y_{ij}} \propto \omega^*_{\mathbf{y}_i,l}\, \phi_N(y_{ij}; \theta^*_{\mathbf{y}_i,l}).$$

(5) Sample $\pi^*_k$, for $k = 1, \ldots, K$, by generating

$$(u^*_k | \ldots) \sim \text{Be}(1 + m^{\mathbf{x}}_k + n^{\mathbf{y}}_k, \alpha + \sum_{s=k+1}^{K} (m^{\mathbf{x}}_s + n^{\mathbf{y}}_s)), \quad k = 1, \ldots, K-1$$

$$u^*_K = 1,$$

where $\text{Be}(\cdot, \cdot)$ denote the Beta density and $m^{\mathbf{x}}_k$ and $m^{\mathbf{y}}_k$ denote the number of pre- and post-treatment distributions assigned to $k$, and setting $\pi^*_k = u^*_k \prod_{s=1}^{k-1}(1 - u^*_s)$.

(6) Sample $\omega^*_{kl}$ for $k = 1, \ldots, K$ and $l = 1, \ldots, L$ by generating

$$(v^*_{kl} | \ldots) \sim \text{Be}(1 + n^{\mathbf{x}}_{kl} + n^{\mathbf{y}}_{kl}, \gamma + \sum_{s=l+1}^{L} (n^{\mathbf{x}}_{ks} + n^{\mathbf{y}}_{ks})), \quad l = 1, \ldots, L-1$$

$$v^*_{Lk} = 1,$$

where $n^{\mathbf{x}}_{kl}$ and $n^{\mathbf{y}}_{kl}$ denote respectively the number of observations pre- and post-treatment assigned to atom $l$ of distribution $k$, and setting $\omega^*_{kl} = v^*_{kl} \prod_{s=1}^{L-1}(1 - v^*_{ks})$.

(7) Sample $\mu^*_{kl}$ and $\sigma^*_{kl}$ by generating them from Normal-Inverse Gamma distribution with updated parameters:

$$(\mu^*_{kl}, \sigma^*_{kl}) \sim NIG(\mu^{**}, k^{**}, a^{**}, b^{**})$$

with

$$\mu^{**} = \frac{k_0 \mu_0 + \bar{\mu}_{obs}}{k_0 + n^*_{kl}},$$

$$k^{**} = k_0 + n^*_{kl}$$

$$a^{**} = a_0 + n^*_{kl}/2$$

$$d^{**} = d_0 + 0.5 \sum_{i,j | \xi_{\mathbf{x}_i} = k, \xi_{\mathbf{y}_i} = k, \psi_{X_{ij}} = l, \psi_{Y_{ij}} = l} (x_{ij} + y_{ij} - \bar{\mu}_{obs})^2 + \frac{k_0 n^*_{kl} (\bar{\mu}_{obs} - \mu_0)^2}{2(k_0 + n^*_{kl})},$$

where $n^*_{kl}$ is the number of observations both pre- and post- treatment assigned to atom $l$ of distribution $k$ and $\bar{\mu}_{obs}$ is their average.

(8) Sample $\alpha$ and $\gamma$ by generating them from a Gamma distribution with updated parameters:

$$\alpha \sim Gam(a_\alpha + (K-1), b_\alpha - \sum_{k=1}^{K-1} \log(1-u_k^*)),$$

$$\gamma \sim Gam(a_\gamma + K(L-1), b_\gamma - \sum_{l=1}^{L-1}\sum_{k=1}^{K-1} \log(1-v_{kl}^*)).$$

**Table 1**

*Marginal Regressions of individual pre- and post- treatment mean values on hemoglobin levels and increase in prostate antigen levels to test the validity of the simplifying assumption of no association between the distributions of the PDGFR values $X$ and $Y$ and the other covariates $Z$ in the PFS model*

| | $\hat{\beta}_0$ (p-value) | $\hat{\beta}_1$ (p-value) |
|---|---|---|
| $\overline{X} = \beta_0 + \beta_1 PSA$ | 13.70 (<0.001) | 0.0004 (0.241) |
| $\overline{Y} = \beta_0 + \beta_1 PSA$ | 13.48 (<0.001) | -0.0004 (0.417) |
| $\overline{X} = \beta_0 + \beta_1 Hem$ | 13.75 (< 0.001) | -0.068 (0.704) |
| $Y = \beta_0 + \beta_1 Hem$ | 13.56 (<0.001) | 0.088 (0.767) |

**Table 2**

*Fit of extended Cox model studying the time-varying effect of a shift in PDGFR values on progression free survival. The covariate $X$ is PDFGR inhibition, coded 0 if $E(\Delta|data) < 0.5$ and 1 if $E(\Delta|data) \geqslant 0.5$. The model has hazard function $h(t|X) = h_0(t)\exp(\beta_1 X + \beta_2 X * log(t))$. Similar time-varying models provided analogous results. $\hat{\beta}$ indicates the estimated regression coefficient.*

| | $\hat{\beta}$ | $\exp\hat{\beta}$ | $\text{se}(\hat{\beta})$ | $z$ | $p$-value |
|---|---|---|---|---|---|
| $X$ | 2.98 | 19.748 | 0.937 | 3.18 | 0.0015 |
| $X\log(t)$ | -1.54 | 0.214 | 0.599 | -2.58 | 0.0100 |

**Table 3**
*Posterior probability of observing a value of $\Delta$ smaller, equal or greater than 0.5, in the two arms of the study described in Section 7.*

| | $P(\Delta < 0.5 \mid \text{data})$ | $P(\Delta = 0.5 \mid \text{data})$ | $P(\Delta > 0.5 \mid \text{data})$ |
|---|---|---|---|
| $DI$ arm | 0.34 | 0.34 | 0.32 |
| $D$ arm | 0.19 | 0.45 | 0.36 |