Biometrics 71, 188–197
March 2015

DOI: 10.1111/biom.12250

Bayesian Nonparametric Estimation of Targeted Agent Effects on Biomarker Change to Predict Clinical Outcome

Rebecca Graziani,[1] Michele Guindani,[2,*] and Peter F. Thall[2]

[1]Università Commerciale Luigi Bocconi, Milan, Italy
[2]University of Texas MD Anderson Cancer Center, Houston, Texas, U.S.A.
[*]*email:* mguindani@mdanderson.org

Summary. The effect of a targeted agent on a cancer patient's clinical outcome putatively is mediated through the agent's effect on one or more early biological events. This is motivated by pre-clinical experiments with cells or animals that identify such events, represented by binary or quantitative biomarkers. When evaluating targeted agents in humans, central questions are whether the distribution of a targeted biomarker changes following treatment, the nature and magnitude of this change, and whether it is associated with clinical outcome. Major difficulties in estimating these effects are that a biomarker's distribution may be complex, vary substantially between patients, and have complicated relationships with clinical outcomes. We present a probabilistically coherent framework for modeling and estimation in this setting, including a hierarchical Bayesian nonparametric mixture model for biomarkers that we use to define a functional profile of pre-versus-post-treatment biomarker distribution change. The functional is similar to the receiver operating characteristic used in diagnostic testing. The hierarchical model yields clusters of individual patient biomarker profile functionals, and we use the profile as a covariate in a regression model for clinical outcome. The methodology is illustrated by analysis of a dataset from a clinical trial in prostate cancer using imatinib to target platelet-derived growth factor, with the clinical aim to improve progression-free survival time.

Key words: Bayesian nonparametrics; Biomarker profiles; Nested Dirichlet process; Receiving operating curve; Survival analysis.

1. Introduction

Intensive research investigating the biological bases for many types of cancer has produced a rapidly growing number of so-called "targeted agents." Targeted agents may be molecules, such as tyrosine kinase inhibitors, or cells, such as T-cells or natural killer cells engineered to recognize a particular cell surface marker. The rationale for using a targeted agent therapeutically is that it causes either direct destruction of the cancer cell or a change in a pathway believed to be associated with an observable change in a biomarker related to clinical outcome. The therapeutic goal may be to stabilize a solid tumor so that it stops growing, reduce the patient's disease burden or, ideally, eradicate the disease. Identification of biomarkers that are changed by a targeted agent is a major challenge due to the complexities of biological disease mechanisms, which make it difficult to estimate effects of targeted agents on the levels of biomarkers and their associated pathways (Ratain and Glassman, 2007). Once a biomarker affected by a targeted agent has been identified, establishing a relationship between the biomarker and clinical outcome also is difficult due to variability in the biomarker, evaluation methods, and patient response. See, for example, the recent discussion in Kelloff and Sigman (2012).

In the following formalism, each object may be one-dimensional or a vector of several elements. Denote treatment regime by τ, and available covariates by Z. Let X denote the baseline biomarker and Y its corresponding post-treatment value, and denote clinical outcome by T. In practice, τ may be a single agent, a combination or sequence of two or more agents, or an administration mode. Let $p(\cdot)$ denote a distribution, for example, $p(T \mid X, Y, Z, \tau)$ is the conditional distribution of T given X, Y, Z, τ.

Putatively, a desired effect of a targeted τ on T is mediated, at least in part, through the effect of τ on the biomarker, in particular the change from X to Y. This effect first is identified in pre-clinical experiments, which provide the rationale for developing and using targeted agents therapeutically on humans. Once τ has been used to treat humans, a first question is how the biomarker distribution may be affected by τ. Addressing this requires characterizing and estimating the change within each patient from X to Y. To assess this change, one must account for association between X and Y due to within-patient effects, in addition to treatment effects on Y and covariate effects on both X and Y, hence $p(Y \mid X, Z, \tau)$ and $p(X \mid Z)$. However, these distributions may be complex and highly disperse. In many settings, X and Y are dichotomized, for example, to identify nominally low versus high biomarker expression. This simplification may misrepresent the data by discarding important information, especially quantitative within-patient biomarker changes due to treatment. Discussions of information loss or distortion due to discretizing continuous variables are given by Altman et al. (1994), Irwin and McClelland (2003), and Royston, Altman, and Sauerbrei (2006). Recent studies have reported multimodal or skewed distributions of putative biomarkers (e.g., Lucas et al., 2009; Bessarabova et al., 2010), and some authors have proposed indexes of bimodality for scoring transcript expression profiles (Wan et al., 2009). If $p(X \mid Z)$

and $p(Y \mid X, Z, \tau)$ are multi-modal and highly complex, empirical estimates of quantities like $\mathrm{E}(X - Y \mid Z, \tau)$ may not be meaningful (cf. Morita et al., 2010).

In this article, we develop a unified, probabilistically coherent framework for quantifying treatment effects on a given biomarker's levels and the effect of such changes on T. We present a comprehensive model including a *biomarker functional profile*, $\Delta = \Delta\{p(X, Y \mid Z, \tau)\}$, to characterize biomarker changes due to treatment, and we integrate this into a submodel including the effect of Δ on T. The model also can identify subgroups of individuals characterized by different Δ profiles in terms of the effects of τ and Δ on T, which is useful for prognostic purposes. More specifically, we employ the probability factorization,

$$p(T, X, Y \mid Z, \tau) \;=\; p(T \mid X, Y, Z, \tau)\, p(X, Y \mid Z, \tau). \qquad (1)$$

In (1), we will make two simplifying assumptions. The first is that $p(T \mid X, Y, Z, \tau) = p(T \mid \Delta, Z, \tau)$, that is, given Z and τ the effect of (X, Y) on T is characterized by Δ. Although in general the distributions of X and Y may depend on Z, for simplicity hereafter we also will assume that neither X nor Y depends on Z, so that $p(X, Y \mid Z, \tau) = p(X, Y \mid \tau)$. We employ a Bayesian parametric regression model for $p(T \mid \Delta, Z, \tau)$, and a hierarchical Bayesian nonparametric model for $p(X, Y \mid \tau)$, which gives flexible subject-specific estimates of Δ, and clusters of these estimates.

Our motivating application is a dataset from a randomized clinical trial in men with advanced prostate cancer (Matthew et al., 2007; Morita et al., 2010). The goal was to compare two drug combinations, docetaxel plus imatinib (DI, $n = 41$) and docetaxel plus placebo (D, $n = 47$). Docetaxel was an established chemotherapy for this disease, while imatinib had shown substantial beneficial effects in other diseases, most notably chronic myelogenous leukemia, and, based on preclinical experiments, also was considered promising for prostate cancer. The target was the platelet-derived growth factor receptor (PDGFR), which is implicated in cancer cell angiogenesis and believed to increase the chance of the prostate cancer metastasizing to the bone. The hypothesis was that imatinib would reduce the blood concentrations of phosphorylated PDGFR (p-PDGFR), thus inhibiting tumor angiogenesis, reducing the development of bone metastases, and improving progression free survival (PFS) time. The dataset includes paired samples of p-PDGFR values from peripheral blood leukocytes, obtained from each patient at baseline and after chemotherapy. Additional covariates included hemoglobin and prostate-specific antigen (PSA) levels.

In order to obtain a highly flexible yet tractable model for $p(X, Y \mid \tau)$, we use the Nested Dirichlet Process (NDP) (Rodriguez, Dunson, and Gelfand, 2008), a hierarchical Bayesian nonparametric model that has been applied extensively for density estimation. In the present setting, the NDP model first assumes that each patient's observed X and Y are conditionally independent samples from mixtures of normal distributions having parameters assumed to follow priors that are patient-specific realizations of Dirichlet processes. A hierarchical structure is obtained by assuming that the patient-specific Dirichlet processes are conditionally independent samples from a hyperprior (second level prior) that also is a Dirichlet process. This NDP structure accommodates highly complex, multi-modal distributions for the observed vectors of X and Y values of each patient, substantial between-patient variability, identifies patient clusters, and also describes population properties of the biomarkers.

To characterize biomarker change, we propose a functional biomarker profile Δ by building on Bayesian nonparametric density estimation (Ferguson, 1983; Escobar and West, 1995) to obtain an estimator of $P(X < Y \mid \tau)$ that accounts for complex distributional forms of X and Y. Inference about $P(X < Y \mid \tau)$, sometimes referred to as "stress-strength reliability," has received wide attention in the literature (see, e.g., Kotz, Lumelskii, and Pensky, 2006, for a review). Recent work in the Bayesian literature includes Ventura and Racugno (2011), where inference is conducted via pseudo-likelihoods, and Rubio and Steel (2012), where the dependencies between X and Y are modeled parametrically through a Gaussian copula. Our approach is similar to that proposed by Branscum et al. (2008) for disease diagnosis and for quantifying the discriminatory ability of a continuous diagnostic measure. However, our interest resides in evaluating and clustering individual responses in an integrated survival framework, not on the population-level performance of a screening test.

The article is organized as follows. Section 2 describes the general modeling framework. In Section 3, we detail the NDP model for characterizing individual patient profiles. Section 4 discusses the functional profile we use to characterize biomarker distributional change. Section 5 describes posterior computation. In our dataset, we have available large within-patient samples of the baseline biomarker X and the corresponding post-treatment levels Y, as it occurs in many applications involving tissue or blood cell samples. The special case where X and Y are single quantitative, categorical or binary, variables is discussed in Section 6. In Section 7, we apply our method to analyze data from a randomized clinical trial of imatinib in prostate cancer. Section 8 concludes with a brief discussion.

2. Integrated Survival Model

In this Section, we establish notation for the data structure and introduce our model. We index subjects by $i = 1, \ldots, N$. Although our framework easily may accommodate any univariate outcome, in the following we assume that T_i is a time-to-event outcome, such as PFS or overall survival time. Let T_i^o denote either the observed time of the event or right-censoring, with $\varepsilon_i = 1$ if $T_i^o = T_i$ and 0 if $T_i^o < T_i$. Let $\boldsymbol{T}^o = (T_1^o, \ldots, T_N^o)$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$, and let $\mathbf{Z}_i = (Z_{1i}, Z_{2i}, \ldots, Z_{ki})$ denote baseline prognostic covariates, with $\boldsymbol{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N)$. For the ith individual, let n_i and m_i denote the respective frequencies of measurements of the biomarker levels collected before and after treatment. For example, the biomarker expression level may be obtained for each cell in blood samples before and after chemotherapy. Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})$ and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})$ denote the subject-specific pre-treatment and post-treatment biomarker samples, respectively, with $\boldsymbol{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$ and $\boldsymbol{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$.

The functional $\Delta_i = \Delta\{p(\mathbf{X}_i, \mathbf{Y}_i|\tau_i)\}$ characterizes the individual change in the levels of the biomarker before and after

treatment, where τ_i denotes the treatment assigned to the ith individual. For example, Δ may be the L^1-norm distance or some other measure of distributional distance, such as that discussed below in Section 4. Following (1), the patient-level data likelihood can be factored as

$$\begin{aligned} &p(T_i^o, \varepsilon_i, \mathbf{X}_i, \mathbf{Y}_i|\mathbf{Z}_i, \tau_i, \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &\quad = p(T_i^o, \varepsilon_i|\Delta_i, \mathbf{Z}_i, \tau_i, \boldsymbol{\beta})\, p(\mathbf{X}_i, \mathbf{Y}_i|\tau_i, \boldsymbol{\theta}), \end{aligned} \tag{2}$$

where $\boldsymbol{\theta}$ is the vector of parameters for the model on the biomarker profiles and $\boldsymbol{\beta}$ parameterizes the regression model $p(\boldsymbol{T}^o, \boldsymbol{\varepsilon} \mid \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \tau, \boldsymbol{\beta}) = p(\boldsymbol{T}^o, \boldsymbol{\varepsilon}|\Delta, \mathbf{Z}, \tau, \boldsymbol{\beta})$. For example, $\boldsymbol{T}^o$ may follow a commonly used survival distribution such as the Weibull or lognormal, with the linear component being a function of Δ, τ, and $\boldsymbol{Z}$.

3. Bayesian Nonparametric Model for Biomarkers

The model for $p(\mathbf{X}, \mathbf{Y}|\tau, \boldsymbol{\theta})$ must be flexible enough to represent a wide range of possible biomarker distributions that may be multimodal or skewed, and account for variability between individuals. To describe such heterogeneity, we use a hierarchical Bayesian nonparametric framework. The model gives highly flexible individual density estimates that can be used to identify groups (clusters) of individuals characterized by their biomarker profiles. More specifically, we assume that the biomarker measurements are samples from unknown subject-specific distributions with $X_{i1}, \ldots, X_{in_i} \overset{\text{iid}}{\sim} F_{X_i}$, and $Y_{i1}, \ldots, Y_{im_i} \overset{\text{iid}}{\sim} F_{Y_i}$, where X_i and Y_i are the vectors of subject-specific pre-treatment and post-treatment measurements. We model the F_{X_i}'s and F_{Y_i}'s using mixtures of Gaussian distributions.

Denote the Gaussian distribution with mean μ and standard deviation σ by $N(\mu, \sigma)$, and the corresponding pdf and cdf, respectively, by $\phi(\cdot, \mu, \sigma)$ and $\Phi(t; \mu, \sigma)$. Standard Gaussian distributions will be denoted using only the first argument. We first assume that, for each individual $i = 1, \ldots, N$,

$$\begin{aligned} X_{ij}|\mu_{X_{ij}}, \sigma_{X_{ij}} &\overset{\text{ind}}{\sim} N(\mu_{X_{ij}}, \sigma_{X_{ij}}), \quad j = 1, \ldots, n_i, \\ Y_{ik}|\mu_{Y_{ik}}, \sigma_{Y_{ik}} &\overset{\text{ind}}{\sim} N(\mu_{Y_{ik}}, \sigma_{Y_{ik}}), \quad k = 1, \ldots, m_i. \end{aligned} \tag{3}$$

Denote $\theta_{X_{ij}} = (\mu_{X_{ij}}, \sigma_{X_{ij}})'$ and $\theta_{Y_{ij}} = (\mu_{Y_{ij}}, \sigma_{Y_{ij}})'$. Under the mixture model, $\theta_{X_{ij}}$ and $\theta_{Y_{ij}}$ are sampled from some mixing distributions, G_{X_i} and G_{Y_i}, that is,

$$\theta_{X_{i1}}, \ldots, \theta_{X_{in_i}}|G_{X_i} \overset{\text{iid}}{\sim} G_{X_i} \text{ and } \theta_{Y_{i1}}, \ldots, \theta_{Y_{im_i}}|G_{Y_i} \overset{\text{iid}}{\sim} G_{Y_i}. \tag{4}$$

This implies that, conditional on realizations G_{X_i} and G_{Y_i}, the distributions of the individual vectors $\mathbf{X}_i$ and $\mathbf{Y}_i$ are the mixtures,

$$\begin{aligned} f_X(\mathbf{x}_i|G_{X_i}) &= \int \prod_{j=1}^{n_i} \phi(x_{ij}; \theta_{X_{ij}})\, G_{X_i}(\mathrm{d}\theta_{X_{ij}}), \\ f_Y(\mathbf{y}_i|G_{Y_i}) &= \int \prod_{k=1}^{m_i} \phi(y_{ik}; \theta_{Y_{ik}})\, G_{Y_i}(\mathrm{d}\theta_{Y_{ik}}). \end{aligned} \tag{5}$$

Here, our focus is on assessing the change in the distribution of $\mathbf{Y}_i$ versus $\mathbf{X}_i$, in terms of Δ_i, to determine the effect of a given treatment, investigate possible associations of the change with clinical outcomes, and identify groups (clusters) of individuals having similar biological responses. This requires a prior model on G_{X_i} and G_{Y_i} that identifies clusters of individual profiles. For that purpose, our prior model on G_{X_i} and G_{Y_i} involves the Dirichlet Process (DP), a widely used prior probability model for unknown distributions, which often is used for its clustering properties (see Ferguson, 1973). We write $G \sim \mathrm{DP}(\alpha, G^\star)$ to indicate that a random distribution G follows a DP with base measure $E(G) = G^\star$ and total mass parameter α. The parameter α determines, among other important properties, the variation of G around the prior mean, with smaller (larger) values of α implying higher (lower) uncertainty. A random probability measure G with DP prior a.s. has discrete support, and can be represented as $G = \sum_{h=1}^{\infty} \pi_h \delta_{m_h}$, where δ_m is the degenerate distribution with point mass 1 at m_h, with $\pi_h = v_h \prod_{j=1}^{h-1}(1 - v_j)$ for $v_h \sim \mathrm{Be}(\alpha, 1)$ and $m_h \sim G^\star$, $h = 1, \ldots, \infty$ (Sethuraman, 1994). Since G is a.s. discrete, samples from G have a positive probability of ties, as some sets of $\theta_{X_{ij}}$'s ($\theta_{Y_{ij}}$) in (4) may be equal. A review of DP mixture models is given by Hjort et al. (2010).

A DP can be used to flexibly estimate G_{X_i} and G_{Y_i} for each individual, but it does not provide an explicit clustering of those distributions either across i or between pre- and post-treatment samples. Such clustering is important to properly characterize the functional Δ_i. We achieve this objective by assuming that G_{X_i} and G_{Y_i} are realizations of a common NDP (Rodriguez et al., 2008). In this framework, the individual realizations of G_{X_i} and G_{Y_i} can be shared both across individuals and between each individual's pre- and post-treatment samples. Formally, G_{X_i} and G_{Y_i} are conditionally independent samples from a common DP prior, that is,

$$G_{X_i} \sim \sum_{r=1}^{\infty} \pi_r\, \delta_{G_r^*} \quad \text{and} \quad G_{Y_i} \sim \sum_{r=1}^{\infty} \pi_r\, \delta_{G_r^*}, \tag{6}$$

where each atom G_r^* is itself a realization from a $\mathrm{DP}(\gamma, G_0)$ hyperprior with base measure G_0 and concentration parameter γ, that is $G_r^*(\cdot) = \sum_{l=1}^{\infty} w_{lr}^* \delta_{\theta_{lr}^*}(\cdot)$, with $\theta_{lr}^* \sim G_0$. Thus, each G_{X_i} and G_{Y_i} is sampled from a collection of distributions, the G_r^*'s. We denote this hierarchical NDP formulation as $(G_{X_i}, G_{Y_i}) \sim \mathrm{NDP}(\alpha, \gamma, G_0)$. Recall that Δ_i is a functional of $p(\mathbf{X}_i, \mathbf{Y}_i|\mathbf{Z}_i, \tau_i)$, which in the NDP is determined by the realizations of G_{X_i} and G_{Y_i} under (6). Because the DP given by (6) has discrete support, if some π_r's in (6) are not negligible then there is a non-trivial probability that $G_{X_i} = G_{Y_i}$. This implies that the biological effect of treatment is null for patient i, which should be reflected by the posterior estimate of Δ_i. This is discussed in Section 4. For the same reason, there also is non-trivial probability that $(G_{X_i}, G_{Y_i}) = (G_{X_{i'}}, G_{Y_{i'}})$ for some $i \neq i'$, and thus $\Delta_i = \Delta_{i'}$, which says that the biomarker profiles of subjects i and i' belong to the same cluster. The model is completed by specifying the base measure G_0, which we assume follows a Normal-Inverse Gamma (N-IG) prior distribution, as is commonly done for mean and precision vectors in Normal models. Finally, independent Gamma priors are assigned to α and γ.

4. Biomarker Distributional Change Functionals

In this section, we discuss definitions of the biomarker functional profile, Δ. Measures of distributional change based on Bayesian nonparametric priors recently have been considered by Nguyen and Gelfand (2014), who describe L^r-norm, variational and symmetrized Kullback–Leibler distances between realizations of a random probability measure in a functional ANOVA setting. Here, our objective is different, as our goals are to estimate how a biomarker changes due to treatment and relate this to the outcome T. For example, if on average a given treatment results in a decrease of a biomarker's levels, this information usually is hidden by area-based measures such as those noted above. Therefore, we employ a measure of distributional change that more appropriately describes the probability of a shift in the distribution of biomarker levels before and after treatment. In the literature, such a measure has been provided, for example, by the vertical quantile comparison function (see Li, Tiwari, and Wells, 1996, 1999).

For two cdfs F_X and F_Y with common support, the vertical quantile comparison function is defined as $Q_{X,Y}(p) = F_Y\{F_X^{-1}(p)\}$, for $p \in (0, 1)$. This quantity represents the functional form of the probability plot and facilitates comparison of two distributions' quantiles. In particular, $Q_{X,Y}(p) = 0.5$ has been used in so-called "control median tests" for comparing control and treatment groups (see Gart, 1963; Gastwirth, 1968; Park, 2002). In decision theory, the vertical quantile comparison function is related to the Receiving Operating Characteristic (ROC) curve, defined as $\mathrm{ROC}(p) = 1 - F_Y\{F_X^{-1}(1-p)\}$. In that context, F_X and F_Y represent the cdfs of a diagnostic variable in two populations, for example, healthy controls and patients, respectively. See also the use of placement values in diagnostic testing (Pepe and Cai, 2004). Here, we are not interested in assessing the diagnostic performance of a biomarker. Nevertheless, in the evaluation of targeted therapies, the vertical quantile comparison function can be estimated by considering the distribution functions F_{X_i} and F_{Y_i} of the individual biomarker levels before and after treatment. We define a measure of distributional change as $\Delta = \int_0^1 Q_{X,Y}(p)\mathrm{d}p = E_{F_Y}\{F_X(Y)\} = P(X < Y)$, which corresponds to the area under the ROC curve commonly used in diagnostic testing.

In the present setting, Δ_i represents the shift in biomarker distribution of the ith subject, putatively due to the treatment's biological mechanisms of action. A posterior estimate with $P(X_{ij} < Y_{ik} \mid \text{data}) > 0.5$ implies that the subject's distribution has shifted to the right (biomarker increase), whereas $P(X_{ij} < Y_{ik} \mid \text{data}) < 0.5$ implies a shift to the left (biomarker decrease), and $P(X_{ij} < Y_{ik} \mid \text{data}) \approx 0.5$ corresponds to no substantial change in the subject's biomarker levels.

5. Posterior Computation

Figure 1 summarizes the formulation of our hierarchical model. Since the joint posterior distribution of the model parameters cannot be obtained in closed form, we employ Markov chain Monte Carlo (MCMC) algorithms for posterior inference. The full-conditionals for the update of the NDP mixture model parameters are obtained as in Rodriguez et al. (2008), based on a truncation of the Dirichlet Processes. The parameters of the baseline distribution G_0 are updated at each iteration using all samples of biomarker levels, $(\mathcal{X}, \mathcal{Y})$. The full-conditionals for the update of the event time model parameters are obtained using standard methods for Bayesian analysis of accelerated failure survival models. Additional details of the Gibbs sampler are available in the Web Appendix.

To provide a basis for estimating both biological and clinical effects simultaneously, our integrated model fully accounts for the uncertainty of all random quantities, including variability between the Δ_i's to reflect heterogeneity of the patient population. At each iteration, the MCMC algorithm draws samples of the biomarker distributions (G_{X_i}, G_{Y_i}) for each individual, and uses these to obtain an update of the biomarker functional profile Δ_i. As an illustration, consider the model as outlined in Figure 1. To obtain the MCMC posterior samples after burn in, each sample value Δ_i^* is computed by averaging over the posterior estimates $(G_{X_i}^*, G_{Y_i}^*)$ of that subject's biomarker distributions. Since (5) and (6) imply that, for any given mixing distribution G_r^*, $F(t|G_r^*) = \int \Phi(t;\theta)G_r^*(\mathrm{d}\theta) = \sum_{l=1}^{\infty} w_{lr}^*\Phi(t;\theta_{lr}^*)$ for $F = G_{X_i}$ or G_{Y_i},, the posterior biomarker profile is computed as $\Delta_i^* = E_{G_{Y_i}^*}\{G_{X_i}^*(Y_{ik})\} = \int G_{X_i}^*(y)\,\mathrm{d}G_{Y_i}^*(y)\mathrm{d}y$. Denoting $G_{X_i}^* = \sum_{l=1}^{\infty} w_l^*\delta_{\theta_l^*}$ and $G_{Y_i}^* = \sum_{l'=1}^{\infty} w_{l'}^*\delta_{\theta_{l'}^*}$, it follows that the posterior biomarker functional profile estimate is the mean over posterior values of the form $\Delta_i^* = \sum_l \sum_{l'} \omega_l^*\omega_{l'}^* \left(1 - \Phi\left(\frac{\mu_l^* - \mu_{l'}^*}{\sqrt{\sigma_l^{2*} + \sigma_{l'}^{2*}}}\right)\right)$.

6. Binary or Single Biomarker Measurements

In many settings, the biomarker is binary, either as the result of dichotomizing a continuous variable or as an indicator of the presence/absence of a biological characteristic in the patient, such as a gene mutation. It also commonly is the case that biomarker values are available only as pre–post-pairs of single measurements on each patient. Additionally, clinical outcome may not be an event time but rather a categorical variable, such as {complete response, partial response, stable disease, progressive disease}, or a binary variable such as the indicator of complete or partial response. In this section, we show how our modeling framework can accommodate such settings.

For the case where the elements of $\mathbf{X}_i$ and $\mathbf{Y}_i$ are binary, the model structure established above may be exploited by assuming that it holds for unobserved real-valued latent variables, say $\tilde{X}_{ij}$ and $\tilde{Y}_{ik}$, and defining the observed binary variables as $X_{ij} = \mathrm{I}(\tilde{X}_{ij} > 0)$ and $Y_{ij} = \mathrm{I}(\tilde{Y}_{ik} > 0)$. This follows the latent variable data augmentation approach of Albert and Chib (1993). One then can think of X_{ij} and Y_{ik} as draws from conditionally independent Bernoulli distributions with subject-specific parameters $p_{X_{ij}}$ and $p_{Y_{ik}}$, that is, $X_{ij} \sim \mathrm{Bern}(p_{X_{ij}})$ and $Y_{ik} \sim \mathrm{Bern}(p_{Y_{ik}})$. These probabilities are then mapped to real-valued parameters, $\theta_{X_{ij}} = \mathrm{link}(p_{X_{ij}})$, $\theta_{Y_{ik}} = \mathrm{link}(p_{Y_{ik}})$, by a conventional link function such as the logit or probit. The established NDP prior construction is then used to obtain individual estimates of Δ_i, as above. However, note that here $\Delta_i = \Pr(X_i = 0, Y_i = 1)$, and G_{X_i} and G_{Y_i} are mixtures of Bernoulli distributions, allowing the possibility that the biomarker distributions are not the same across samples.

Next, we consider the case where only a single pair of pre- and post-treatment biomarker measurements, (X_i, Y_i), are available for each patient, that is, all $n_i = m_i = 1$. While

Event Time Likelihood

$$T_i \mid \boldsymbol{Z}_i, \tau_i, \Delta_i, \beta, \sigma_T \sim f(t_i|\boldsymbol{Z}_i, \tau_i, \Delta_i, \beta, \sigma_T) \quad i = 1, \ldots, N,$$

$$\textit{Example}: \quad \log(T_i) \mid \eta_i(\boldsymbol{Z}_i, \tau_i, \Delta_i, \beta), \sigma_T \sim \mathrm{N}(\eta_i,\ \sigma_T)$$

Biomarker Profiles Likelihood

$$\begin{aligned} X_{ij} \mid \mu_{X_{ij}}, \sigma_{X_{ij}} &\sim \mathrm{N}(\mu_{X_{ij}}, \sigma_{X_{ij}}), \quad j = 1, \ldots, n_i, \\ Y_{ik} \mid \mu_{Y_{ik}}, \sigma_{Y_{ik}} &\sim \mathrm{N}(\mu_{Y_{ik}}, \sigma_{Y_{ik}}), \quad k = 1, \ldots, m_i, \\ \Delta_i &= E_{G_{Y_i}}\{G_{X_i}(Y_{ik})\} = P(X_{ij} < Y_{ik}) \end{aligned}$$

Priors

$$\theta_{X_{ij}} = (\mu_{X_{ij}}, \sigma_{X_{ij}})^T \text{ and } \theta_{Y_{ik}} = (\mu_{Y_{ik}}, \sigma_{Y_{ik}})^T$$
$$\theta_{X_{ij}} \mid G_{X_i} \sim G_{X_i}, \text{ and } \theta_{Y_{ik}}|G_{Y_i} \sim G_{Y_i},$$
$$(G_{X_i}, G_{Y_i}) \sim \mathrm{NDP}(\alpha, \gamma, G_0)$$

$$\alpha \sim Gam(a_\alpha, b_\alpha), \quad \gamma \sim Gam(a_\gamma, b_\gamma)$$
$$G_0 \sim \mathrm{N-IG}(\mu_0, k_0, a_0, d_0)$$
$$(\beta, \sigma_T) \sim \mathrm{N-IG}(\mu_1, k_1, a_1, d_1)$$

Fixed Hyperparameters

$$\mu_0, k_0, a_0, d_0, \mu_1, k_1, a_1, d_1$$

Figure 1. Hierarchical formulation of the proposed probabilistic model. T_i represents the outcome of interest, τ_i is the treatment, $\mathbf{Z}_i$ is a vector of covariates, $\mathbf{X}$ and $\mathbf{Y}$ are the pre- and post-treatment values of a biomarker for patient i.

the two biomarker distributions for each individual now cannot be estimated, population level estimates of biomarker change due to treatment can be obtained. We collapse the hierarchical structure, as follows. For a continuous biomarker, writing $\theta_{X_i} = (\mu_{X_i}, \sigma_{X_i})$ and $\theta_{Y_i} = (\mu_{Y_i}, \sigma_{Y_i})$, we assume $X_i|\theta_{X_i} \sim N(\theta_{X_i})$ and $Y_i|\theta_{Y_i} \sim N(\theta_{Y_i})$, with $\theta_{X_i}|G_X \sim G_X$ and $\theta_{Y_i}|G_Y \sim G_Y$, and $(G_X, G_Y) \sim DP(\alpha, G_0)$. The binary case can be handled similarly, assuming $X_i \sim \mathrm{Bern}(p_{X_i})$ and $Y_i \sim \mathrm{Bern}(p_{Y_i})$ where the individual real-valued parameters $\theta_{X_i} = \mathrm{link}(p_{X_i})$ and $\theta_{Y_i} = \mathrm{link}(p_{Y_i})$ follow the single level DP given above. Finally, although in the application described in the next section we assume a log-normal distribution for the event times, $p(T_i|\mathbf{Z}_i, \Delta_i, \tau_i)$ can be quite general, to accommodate a binary or categorical outcome T_i by assuming an appropriate generalized linear model within the Bayesian framework.

7. Analysis of the Prostate Cancer Trial Data

7.1. *Data and PFS Model*

In this section, we apply our integrated model to analyze the prostate cancer dataset (Matthew et al., 2007). Figure 2 gives the empirical distributions of the log-transformed p-PDGFR sample values pre- and post-treatment for three patients in the DI treatment arm. The histograms illustrate the multi-modality and pronounced left skewness of the p-PDGFR distributions, and the fact that the treatment

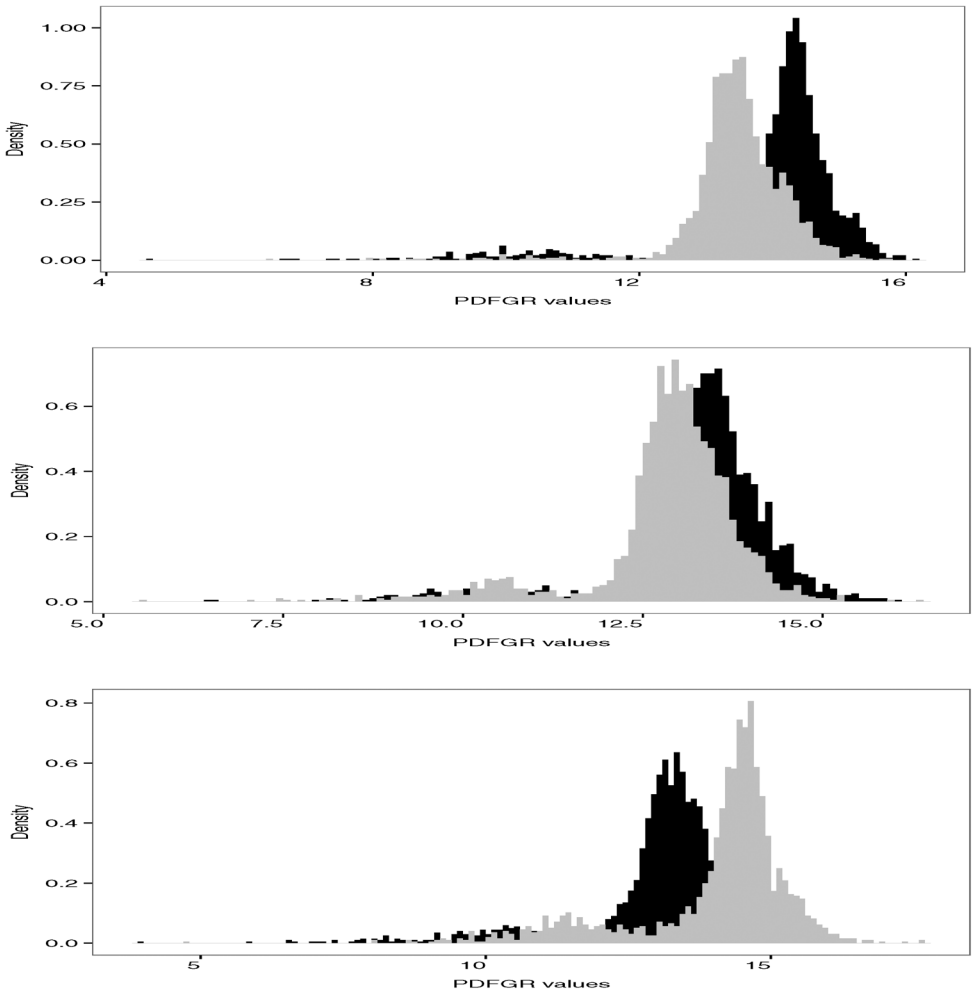

Figure 2. Empirical distributions of log-transformed p-PDGFR values before (black) and after (gray) treatment with docetaxel + imatinib, for three representative patients. The plots highlight the characteristic multi-modality and skewness of these distributions, as well as moderate shifts of their distribution following treatment.

effect was not homogeneous across patients. Compared to $p(X|\tau)$, the post-treatment distribution $p(Y|\tau)$ shifted to the right for some patients, shifted to the left for others, and showed no substantive change for a third subgroup.

Morita et al. (2010) analyzed this dataset using a fully Bayesian parametric hierarchical framework, where both biomarker distributions were modeled as two-component mixtures of normals. They used the differences between the estimated means of the right pre- and post-treatment mixture components, and differences between the estimated means of the left pre- and post-treatment mixture components, as covariates in a parametric regression model for PFS time. This analysis characterizes treatment effect on p-PDGFR by the two estimated location shifts, but it makes no use of any other characteristics of the biomarker distributions. Our Bayesian Nonparametric approach refines this previous analysis in two substantive ways. First, some patients' p-PDGFR distributions exhibited three, four, or five modes, which the previous two-component mixture model ignores but the hierarchical NDP mixture model identifies quite naturally. Second, the NDP model yields an automatic clustering of individuals on the basis of the posteriors of their Δ_i's.

Following Morita et al. (2010), we assume that the T_i's are lognormally distributed, and model the linear term similarly to their formulation. This will facilitate comparison of the two analyses. Let $\tau_i = 0$ if patient i was randomized to the control (D) arm, and $\tau_i = 1$ for the experimental (DI) arm. For the ith patient, define $Z_{1i} = 1$ if the hemoglobin level was greater than 11 g/dl, 0 otherwise, Z_{2i} the pretreatment to post-treatment increase in prostate-specific antigen (PSA) levels, and denote the vector $\mathbf{Z_i}' = (Z_{1i}, Z_{2i})$. We assume that $\log(T_i) \sim \mathrm{N}(\eta_i, \sigma_T)$, where

$$\eta_i = \beta_0 + \beta_1 \tau_i + \left\{\beta_2 \tau_i + \beta_3 (1 - \tau_i)\right\} Z_{1i} + \left\{\beta_4 \tau_i + \beta_5 (1 - \tau_i)\right\} Z_{2i} + \left\{\beta_6 \tau_i + \beta_7 (1 - \tau_i)\right\} \Delta_i. \tag{7}$$

While $\beta_1, \ldots, \beta_7$ have the same respective interpretations as given in Morita et al. (2010), we obtain Δ_i as the area under

the ROC curve, as described in Section 4, based on the NDP model for $\mathbf{X}$ and $\mathbf{Y}$. To test the validity of the simplifying assumption of no association between the distributions of X and Y and the covariates Z, we regressed the individual pre- and post- treatment mean values on hemoglobin levels and increase in prostate antigen levels. This analysis showed that there was no significant association (Web Appendix).

7.2. *Prior Hyperparameter Specification*

Following Ishwaran and James (2001) and Rodriguez et al. (2008), we assume finite truncations of the DP, with the mixture distributions for modeling the pre-treatment and post-treatment biomarker levels in (6) given by G_{X_i} and $G_{Y_i} \sim \sum_{r=1}^{30} \pi_r \delta_{G_r^*}$. Similarly, we define $G_r^*(\cdot) = \sum_{l=1}^{50} w_{lr}^* \delta_{\theta_{lr}^*}(\cdot)$. It is known that this sort of truncation may introduce bias into the results (Griffin and Walker, 2011). To partially deal with this, alternatively retrospective sampling (Papaspiliopoulos and Roberts, 2008) or slice sampling (Walker, 2007; Kalli, Griffin, and Walker, 2011) techniques could be used. However, a preliminary sensitivity analysis considering truncation levels ranging from 20 to 60 produced no significant differences in the final results.

A priori, we assume that the means of each biomarker cluster, $\mu_{X_{ij}}$ and $\mu_{Y_{ij}}$, are drawn from a vague normal distribution centered at 0 with large variance. Specifically, the baseline of the NDP mixture model was set as $G_0 \equiv$ N-IG(0,1,5,50), which implies that $E(\mu|\sigma) = 0$, $\text{SD}(\mu|\sigma) = \sigma$, $E(\sigma) = 12$ and $\text{Var}(\sigma) = 52$. These values were chosen to obtain disperse prior predictive distributions of the X_{ij}'s and Y_{ij}'s having support that includes the domain [0, 30] where all of the observed values occur. This also ensures that the induced priors on the cluster variances, $\sigma^2_{X_{ij}}$ and $\sigma^2_{Y_{ij}}$, are vague. For the precision parameters of the nested DPs, α and γ, we specified a Gamma distribution, Ga(1, 1) (Escobar and West, 1995). For the regression model on PFS time, we assume non informative flat priors on the regression coefficients β and the standard deviation σ. Specifically, the β coefficients are independent a priori and, conditionally on σ, are assigned a diffuse Gaussian prior centered at zero. The sampling variance σ^2 is assumed to have a diffuse Inverse-Gamma prior centered at 1 and infinite variance. This results in a joint multivariate Normal-Inverse Gamma prior specification on the set of β's and σ^2, that is, N-IG($0\,\mathbf{1}_8, \mathbf{I}_8, 2, 1$), where $\mathbf{1}_8$ and $\mathbf{I}_8$ denote the 8-dimensional vector of ones and the 8-dimensional identity matrix, respectively.

7.3. *Results*

The following results are based on MCMC samples of size 10,000, obtained after a burn-in of 10,000 iterations. We assessed the convergence of the MCMC chains by visual inspection and, more formally, by Raftery and Lewis's diagnostic test and other tests implemented in the R package "coda" (Raftery and Lewis, 1992; Plummer et al., 2006). As expected, the Nested-Dirichlet Process Mixture model confirms the existence of heterogeneous biomarker profiles in patients. Indeed, posterior inference based on the MCMC iterations suggests that the patients' posterior distributions can be clustered approximately into 10 groups (modal value) either before and after treatment (Figure 3). Furthermore, if we consider the distribution of PDGFR values, the number

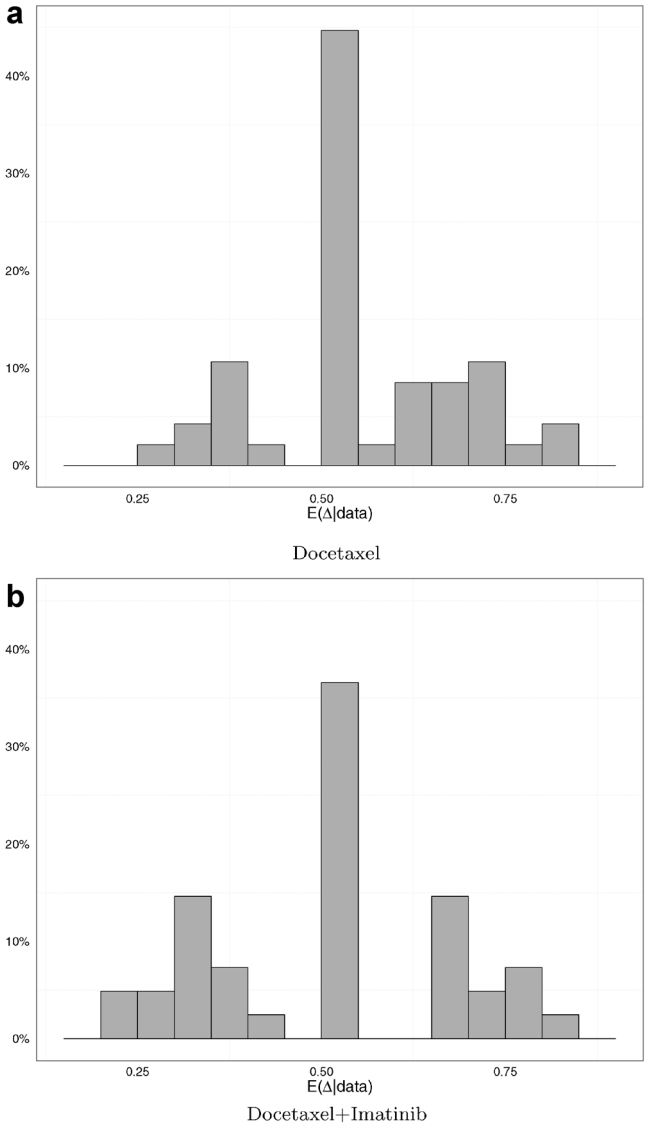

Figure 3. Clustered posterior means of Δ, obtained by averaging each patient values across MCMC iterations, in each of the two treatment arms.

of mixture components estimated for each patient before and after treatment typically ranges between 2 and 5. Assuming a quadratic loss function, we can estimate the probability of a biomarker shift, Δ, for each individual using the posterior mean, $E(\Delta_i|\text{data})$, obtained by averaging the values computed at each MCMC iteration. Figure 3 shows the resulting frequency distribution of the posterior means for all individuals. It is evident that the model captures existing heterogeneity across patients, as measured by the shift of the biomarker distribution before and after treatment. We identified three clusters of patients, according to the value of Δ_i being less than, equal or greater than 0.5. The high percentage of patients for whom there was no significant shift in the distribution of biomarker profiles before and after treatment suggests that the leading hypothesis of the study, that is, that Imatinib lowers the expression of PDGFR values, may not

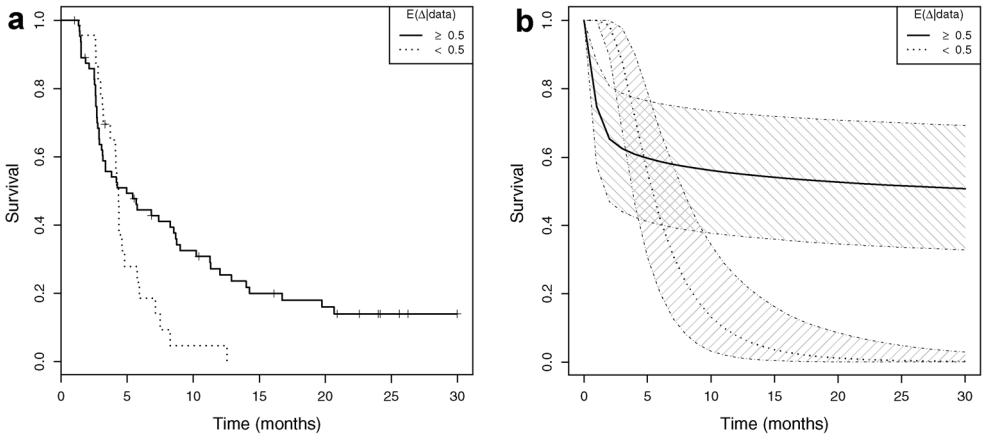

Figure 4. On the left, Kaplan–Meier estimates of progression free survival for patients experiencing a shift to the left (dashed line) or to the right (solid line) in PDFGR values. On the right, posterior median survival and 95% HPD intervals for two representative individuals in the two groups, respectively.

be correct. In fact, the clinical trial was stopped early due to futility, since the initial data suggested no therapeutic benefit on PFS time from the combination of docetaxel and imatinib. 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. See also Table 3 in the Web Appendix. Therefore, contrary to what was initially expected by the investigators, the combination of docetaxel and imatinib did not lead to a large decrease of PDGFR values in the patients' blood. Furthermore, if we consider the study population as a whole and compare their PFS times as functions of the posterior means of Δ_i (probability of a shift being less than versus greater than or equal to 0.5), we see a strong indication that increased survival, surprisingly, was associated with larger values of Δ_i. Figure 4a plots the Kaplan–Meier curves of patients experiencing a shift to the left ($E(\Delta|\text{data}) < 0.5$) versus those experiencing a shift to the right ($E(\Delta|\text{data}) \geq 0.5$) in biomarker levels. Figure 4b provides the posterior median survival and 95% highest posterior density (HPD) intervals of two representative individuals in each group. 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 (Web Appendix).

Table 1 displays the estimates of the coefficients in model (7). In addition to the posterior means of the β_j's, we report 95% posterior credible intervals as well as the posterior probability of observing a value of β greater than 0. With the exception of the effect of hemoglobin in the control arm, most credible intervals overlap zero. Therefore, we should be extremely careful in interpreting the direction of the effects estimated by the model. On the other hand, examination of Prob($\beta > 0$|data) seems to confirm the previous conclusions: higher values of Δ, that is, higher values of PDGFR after treatment, tend to be associated with longer PFS time. We should note that the results in Table 1 are similar to those already given in Morita et al. (2010). However, their parametric analysis did not provide clustering of patients into subgroups induced by the distributional changes of the biomarker profiles before and after treatment. The advantage of the nonparametric approach is apparent from the improved goodness of fit, as measured by the achieved L measure statistics under the two approaches (Ibrahim, Chen, and Sinha, 2001). The L measure defines a criterion for model comparison based on minimizing the posterior predictive loss, with lower values indicating a better fit (Gelfand and Gosh, 1998). Table 2 reports the values of the L statistics for both the fit of the density estimates of the pre- and post- treatment PDFGR values (left column) and the full PFS model (right column) obtained by our approach and the model in Morita et al. (2010). To determine whether the clusters identified by the NDP were predictive of PFS, we also considered an alternative model, where the single predictor Δ was replaced by the indicators of the clusters ($\Delta < 0.5$) and ($\Delta > .0.5$), using ($\Delta = 0$) as the baseline, each interactive with treatment arm, DI or D. The results were very similar to those for model (7). Again, none of the treatment-cluster effects were associated with PFS, with posterior 95% credible intervals of the β_j's for the PDGFR shift overlapping zero.

We also implemented our model assuming that the biomarker values available on each individual are either single measurements or samples of binary indicators. For the first case, we summarized the individual PDGFR values by their

Table 1
Posterior means and 95% *credible intervals of the* β *coefficients in the PFS regression model* (7)

Coefficient	Posterior mean and 95% CI	Prob($\beta > 0$|data)
β_0 (intercept)	1.70 (0.49 2.97)	0.99
β_1 (DI vs. D)	0.35 (−1.04 1.76)	0.66
β_2 (DI hemoglobin)	0.59 (−0.87 1.89)	0.75
β_3 (D hemoglobin)	1.42 (0.02 2.70)	0.97
β_4 (DI PSA)	−0.0014 (−0.016 0.012)	0.42
β_5 (D PSA)	−0.0048 (−0.012 0.002)	0.13
β_6 (DI Δ)	0.30 (−1.15 1.99)	0.63
β_7 (D Δ)	0.84 (−0.61 2.43)	0.79

Table 2
L measure goodness of fit for the density estimates of the pre- and post-treatment PDFGR values (left column) and the full PFS model (right column) obtained by our approach (NDP) and the parametric two-component mixture model in Morita et al. (2010)*. A smaller value of L corresponds to a better fit to the data.*

L measure	Biomarker density	PFS
NDP	5.65	170,300
Parametric	71.93	183,230

means, $\bar{X}_i$ and $\bar{Y}_i$, before and after treatment, respectively, to obtain a single continuous measurement for each individual. To obtain samples of binary measurements, we dichotomized the continuous measurements using the medians as cut-offs. We have applied this dichotomization either to each X_{ij} and Y_{ij}, obtaining a sample of binary biomarker levels for each individual, or to their means, $\bar{X}_i$ and $\bar{Y}_i$, thus obtaining a single binary value of over-expression for each individual, as is commonly done in practice. We then modified our model as discussed in Section 6. The results of the fitted regression models were similar to those in Table 1, although these analyses did not provide additional insights into the heterogeneity of the biomarker profiles across individuals. We have omitted the details of these additional analysis in the interest of space.

8. Discussion

We have proposed an integrated Bayesian hierarchical nonparametric framework for quantifying the effects of treatment on the levels of a biomarker and estimating how such changes affect clinical outcome. The Bayesian nonparametric approach allows us to obtain a flexible estimate of the biomarker profile for each individual and characterize the heterogeneity of patients' responses to treatment. In particular, we have shown that our model can be used to evaluate naive "clinical hypotheses" in subsets of the population of interest. Although our model assumes multiple continuous measurements of a biomarker profile before and after treatment, our framework can be modified to account for binary or single biomarker measurements.

Our application shows how Bayesian hierarchical nonparametric mixture models can help reveal previously unrecognized heterogeneity, and provide a flexible tool for exploring the intricacy of treatment effects on survival. A limitation of the approach presented here is that we assume that the distribution of pre- and post-treatment biomarker values does not depend on covariates. Furthermore, the survival outcome depends on a single measure of functional change, although two different pairs of distributions for X and Y could in principle yield the same $\Delta = P(X < Y)$. Also, we have considered only measurements taken at two time points, before and after treatment. Of course, biomarker levels may change as a complex function of time. If longitudinal biomarker values are available, then a more structured model accounting for patients' biomarker processes would be needed. Extensions of our work may consider dependent bivariate (or multivariate) Dirichlet processes wherein the dependence is obtained by modeling explicitly the relationship between two or more groups of exchangeable data with available covariates (see, e.g., MacEachern, 1999; Walker and Muliere, 2003; De Iorio et al., 2004, 2009; Hatjispyros, Nicoleris, and Walker, 2008). Vector-valued biomarker functional profiles could be considered to capture different features of distributional change (e.g., tail and center behavior).

While our model can only identify groups of patients on the basis of estimated changes in their biomarker levels, further study would be required to link such heterogeneity with the underlying biological mechanisms thought to be responsible for individual effects. Finally, because many targeted agents are not as specific as desired and may cause harm by attacking normal cells, the structure considered here may be extended by considering adverse events, such immunosuppression or organ toxicity, along with *T*.

9. Supplementary Materials

The Web Appendix, referenced in Sections 5 and 7, and the Matlab code implementing the method is available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgements

Peter Thall's work was supported by NCI grant RO1 CA 83932. Peter Thall and Michele Guindani's work was also supported by NCI grant P30-CA016672. The authors would like to thank the Editor, Associate Editor and an anonymous referee for suggestions that substantially improved the article.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Altman, D. G., Lausen, B., Sauerbrei, W., and Schumacher, M. (1994). The dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute* **86**, 829–835.

Bessarabova, M., Kirillow, E., Shi, W., Bugrim, A., Nikolsky, Y., and Nikolskaya, T. (2010). Bimodal gene expression patterns in breast cancer. *BMC Genomics* **11**, S8.

Branscum, A. J., Johnson, W. O., Hanson, T. E., and Gardner, I. A. (2008). Bayesian semiparametric ROC curve estimation and disease diagnosis. *Statistics in Medicine* **27**, 2474–2496.

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics* **65**, 762–771.

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association* **99**, 205–215.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In: *Recent Advances in Statistics*, Rizvi, H., Rustagi, J. (eds), 287–302. New York: Academic Press.

Gart, J. (1963). A median test with sequential applications. *Biometrika* **50**, 55–62.

Gastwirth, J. L. (1968). The first-median test: A two-sided version of the control median test. *Journal of the American Statistical Association* **63**, 692–706.

Gelfand, A. E. and Gosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85**, 1–13.

Griffin, J. E. and Walker, S. G. (2011). Posterior simulation of normalized random measure mixtures. *Journal of Computational and Graphical Statistics* **1**, 241–259.

Hatjispyros, S. J., Nicoleris, T., and Walker, S. G. (2008). Bivariate prior distribution via branching exchangeable sequences. *Journal of Statistical Planning and Inference* **138**, 1799–1816.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge, UK: Cambridge University Press.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). Criterion based methods for Bayesian model assessment. *Statistica Sinica* **11**, 419–443.

Irwin, J. R. and McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research* **40**, 366–371.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of American Statistical Association* **96**, 161–173.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing* **21**, 93–105.

Kelloff, G. J. and Sigman, C. C. (2012). Cancer biomarkers: Selecting the right drug for the right patient. *Nature Reviews Drug Discovery* **11**, 201–214.

Kotz, S., Lumelskii, S., and Pensky, M. (2003). *The Stress-Strength Model and Its Generalizations: Theory and Applications*. New Jersey: World Scientific.

Li, G., Tiwari, R. C., and Wells, M. T. (1996). Quantile comparison functions in two-sample problems with application to comparisons of diagnostic markers. *Journal of the American Statistical Association* **91**, 689–698.

Li, G., Tiwari, R. C., and Wells, M. T. (1999). Semiparametric Inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika* **86**, 487–502.

Lucas, J. E., Carvalho, C. M., Chen, J., Chi, J. T., and West, M. (2009). Cross-study projections of genomic biomarkers: An evaluation in cancer genomics. *PLoS ONE* **4**, e4523.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*.

Mathew, P., Thall, P. F., Bucana, C. D., Oh, W. K., Morris, M. J., Jones, D. M., Johnson, M. M., Wen, S., Pagliaro, L. C., Tannir, N. M., Tu, S. M., Meluch, A. A., Smith, L., Cohen, L., Kim, S. J., Troncoso, P., Fidler, I. J., and Logothetis, C. J. (2007). Platelet-derived growth factor receptor inhibition and chemotherapy for castration-resistant prostate cancer with bone metastases. *Clinical Cancer Research* **13**, 5816–5824.

Morita, S., Thall, P. F., Bekele, B. N., and Mathew, P. (2010). A Bayesian hierarchical mixture model for platelet-derived growth factor receptor phosphorylation to improve estimation of progression-free survival in prostate cancer. *Journal of the Royal Statistical Society, Series C, Applied Statistics* **59**, 19–34.

Nguyen, X. and Gelfand, A. E. (2014). Bayesian nonparametric modeling for functional analysis of variance. *Annals of the Institute of Statistical Mathematics* **66**, 495–526.

Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169–186.

Park, H. I. (2002). Multivariate percentile tests for incomplete data. *Annals of the Institute of Statistical Mathematics* **54**, 934–944.

Pepe, M. S. and Cai, T. (2004). The analysis of placement values for evaluating discriminatory measures. *Biometrics* **60**, 528–535.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11.

Raftery, A. E. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science* **7**, 493–497.

Ratain, M. J. and Glassman, R. H. (2007). Biomarkers in phase I oncology trials: Signal, noise, or expensive distraction? *Clinical Cancer Research* **13**, 6545–6548.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association* **103**, 1131–1154.

Royston, P., Altman, D. G., and Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine* **25**, 127–141.

Rubio, F. J. and Steel, M. F. J. (2012). Bayesian inference for $P(X < Y)$ using asymmetric dependent distributions. *Bayesian Analysis* **7**, 771–792.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Ventura, L. and Racugno, W. (2011). Recent advances on Bayesian inference for $P(X < Y)$. *Bayesian Analysis* **6**, 1–18.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communication in Statistics: Simulation and Computation* **36**, 45–54.

Walker, S. G. and Muliere, P. (2003). A bivariate Dirichlet process. *Statistics and Probability Letters* **64**, 1–7.

Wan, J., Wen, S., Symmans, F., Pusztai, L., and Coombes, K. R. (2009). The bimodality index: A criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Informatics* **7**, 199.

Received October 2013. *Revised August* 2014.
Accepted September 2014.