Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2009 September 3; 59(1): 19–34. doi:10.1111/j.1467-9876.2009.00680.x.

A Bayesian hierarchical mixture model for platelet derived growth factor receptor phosphorylation to improve estimation of progression-free survival in prostate cancer

Satoshi Morita¹, Peter F. Thall², B. Nebiyu Bekele², and Paul Mathew³

Satoshi Morita: smorita@urahp.yokohama-cu.ac.jp

¹Department of Biostatistics and Epidemiology, Yokohama City University Medical Center, Yokohama 232-0024, Japan

²Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77230-1402, U.S.A

³Department of Genitourinary Medical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, U.S.A

SUMMARY

Advances in understanding the biological underpinnings of many cancers have led increasingly to the use of molecularly targeted anti-cancer therapies. Because the platelet-derived growth factor receptor (PDGFR) has been implicated in the progression of prostate cancer bone metastases, it is of great interest to examine possible relationships between PDGFR inhibition and therapeutic outcomes. Here, we analyze the association between change in activated PDGFR (p-PDGFR) and progression free survival (PFS) time based on large within-patient samples of cell-specific p-PDGFR values taken before and after treatment from each of 88 prostate cancer patients. To utilize these paired samples as covariate data in a regression model for PFS time, and because the p-PDGFR distributions are bimodal, we first employ a Bayesian hierarchical mixture model to obtain a deconvolution of the pre-treatment and post-treatment within-patient p-PDGFR distributions. We evaluate fits of the mixture model and a non-mixture model that ignores the bimodality by using a supnorm metric to compare the empirical distribution of each p-PDGFR data set with the corresponding fitted distribution under each model. Our results show that first using the mixture model to account for the bimodality of the within-patient p-PDGFR distributions, and then using the posterior within-patient component mean changes in p-PDGFR so obtained as covariates in the regression model for PFS time provides an improved estimation.

Keywords

Bayesian analysis; Survival analysis; Markov chain Monte Carlo; Platelet derived growth factor receptor; Prostate cancer

1. Introduction

In recent years, clarification of the molecular underpinnings of many types of cancer has yielded improved therapeutic strategies involving so-called “targeted therapies.” If a molecular target is found to be present in a tumor, the strategy with targeted therapy is to achieve anti-disease effect by specific inhibition of the target. Biomarkers associated with therapeutic outcome may facilitate the process of choosing among existing targeted treatments for individual patients.

The platelet-derived growth factor receptor (PDGFR) is an established therapeutic target in several important cancers. PDGFR is of particular interest in men with advanced prostate cancer involving bone metastases because it frequently is over-expressed in this cancer. Denote phosphorylated PDGFR by p-PDGFR. Mathew et al. (2007) conducted a randomized, placebo-controlled clinical trial of the chemotherapeutic agent docetaxel combined with either the oral p-PDGFR inhibitor imatinib mesylate (DI) or placebo (DP) in patients with this disease, with time to disease progression or death (progression free survival, PFS) as the primary endpoint. A key secondary endpoint in this trial was the amount of inhibition of p-PDGFR since the putative effect of imatinib on inhibition of p-PDGFR coupled with the relationship seen between p-PDGFR inhibition in an experimental model of bone metastases (Uehara, et al. 2003) suggested that change in p-PDGFR level from its baseline level pre-treatment to its level post-treatment may be related to PFS time. In order to measure p-PDGFR inhibition, peripheral blood samples were drawn from each patient at two different time points: prior to chemotherapy, and after treatment with one cycle of chemotherapy. In each sample, the intensity of expression of activated p-PDGFR was measured in each of approximately 2000 individual peripheral blood leucocytes using immunofluorescent antibodies, capturing these with laser-scanning cytometry. It was hypothesized that the decrease in p-PDGFR level might be larger in the DI arm due to p-PDGFR inhibition by imatinib and that, in turn, the magnitude of the decrease might be associated with improved PFS. Because very large samples of cell-specific p-PDGFR values were obtained at both measurement times for each patient, very reliable estimates of mean pre-treatment and post-treatment p-PDGFR levels within each patient were available. The difference between the pre- and post-treatment within-patient sample mean p-PDGFR levels proved to be associated with PFS (Mathew et al., 2007). It was noted that in the control arm with docetaxel alone, DP, mean p-PDGFR rose after therapy, but in the DI arm a significantly smaller decrease in p-PDGFR was seen. On average, patients who had a smaller increase in p-PDGFR, corresponding to greater p-PDGFR inhibition, had longer PFS times.

Visual examination of the histograms of the within-patient p-PDGFR samples indicates that their distributions are clearly bimodal, both pre-treatment and post-treatment. This is illustrated by Figure 1, which gives histograms of the log-transformed p-PDGFR values for three typical patients. This observation motivated the statistical re-analysis of this data set that we report here. The main idea of our analysis is to exploit the fact that one may reliably estimate the entire distribution of p-PDGFR expression values at each measurement time for each patient, rather than only their sample means. We hypothesized that accounting for the bimodality of the p-PDGFR distributions would improve the estimation of PFS by providing a more refined representation of p-PDGFR inhibition for use as covariates in the regression model.

Technically, the problem is to estimate each within-patient pre- and post-treatment p-PDGFR distribution, and then choose a small number of features from each distribution for use as covariates in a regression model for PFS. To carry out this analysis, we first fit a Bayesian hierarchical mixture model to the p-PDGFR data accounting for the observed bimodality (McLachlan and Peel, 2000; Gelman, et al., 2004). We compared the fit of this model with that of a corresponding, simpler Bayesian hierarchical model ignoring the bimodality, using a supnorm metric to quantify the distance between the empirical distribution of each observed within-patient p-PDGFR sample and the corresponding fitted distribution obtained under each model. We then used the posteriors of parameters characterizing mean change in p-PDGFR level obtained from either components of the fitted mixture model or, as a basis for comparison, from the fitted non-mixture model, as covariates in a regression model for estimating PFS. In particular, under either hierarchical model for the p-PDGFR data, the posterior mean changes are random quantities, and we

account for this randomness when using these as covariates in the regression models for PFS time. Our results indicate that accounting for the bimodality of the p-PDGFR distributions yields a substantive improvement in the fit of the regression model for PFS.

In Section 2, we describe the data structure of the prostate cancer clinical trial and the Bayesian mixture and survival models. In Section 3, we describe the algorithm that was used to fit the mixture model. We present our analyses of the p-PDGFR data in Section 4, and close with a brief discussion in Section 5.

2. Data structure and models

2.1 Data structure

As mentioned in Section 1, Mathew et al. (2007) conducted a randomized clinical trial to compare DI (docetaxel plus imatinib) and DP (docetaxel plus placebo) in patients with advanced prostate cancer involving bone metastases. They accrued patients between April 2003 and July 2005 at five tertiary cancer care centers in the United States. Paired samples of the p-PDGFR values from peripheral blood leukocytes from 88 men (41 in DI and 47 in DP) were available before and after treatment. The within-patient leukocyte sample sizes $\{m_i\}$ taken pre-treatment and $\{n_i\}$ taken post-treatment were very similar, with median $\{m_i\} = 2021$ (90% ci 2000 – 2074) and median $\{n_i\} = 2031$ (90% ci 2000 – 2103).

For patient $i = 1, \dots, N$, let $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,m_i})$ denote the p-PDGFR values of the m_i cells in the pre-treatment blood sample, $\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,n_i})$ the n_i p-PDGFR values in the post-treatment sample, with $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)$ and $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_N)$. Let T_i denote PFS time, T_i^o the observed value of T_i or right-censoring time and $\varepsilon_i = 1$ if $T_i = T_i^o$ and 0 otherwise. We will utilize the covariates $Z_{1i} = 1$ if patient i received DI and 0 if DP, $Z_{2i} = 1$ if the pre-treatment hemoglobin (Hb) level is greater than 11 g/dl and 0 if not, and Z_{3i} the pre-treatment to post-treatment increase in prostate-specific antigen (PSA) level, and we denote $\mathbf{Z}_i = (Z_{1i}, Z_{2i}, Z_{3i})$ and $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_N)$.

2.2 Non-mixture model for the p-PDGFR data

As a basis for comparison, we first fit the following model that ignores the observed bimodality of the p-PDGFR data. Denote the mean pre-treatment and post-treatment p-PDGFR values for the i^{th} patient by μ_{xi} and μ_{yi} , respectively, with corresponding precision (inverse variance) parameters τ_{xi} and τ_{yi} . Each of these parameters corresponds to the p-PDGFR values of all leukocytes in the patient's peripheral blood at the time the sample was taken. We assume patient-specific means

$\mu_{xi} = \mu_{xi}^* + \xi_i$ and $\mu_{yi} = \mu_{yi}^* + \xi_i$, where $\xi_i = \mu_{xi} - \mu_{xi}^* = \mu_{yi} - \mu_{yi}^*$ is an effect associated with patient i that applies both before and after treatment. Given the patient-specific parameters

$\theta_i = (\xi_i, \mu_{xi}^*, \mu_{yi}^*, \tau_{xi}, \tau_{yi})$, we assume that each pair X_{ij} and Y_{ik} are conditionally independent with the following normal distributions:

$$X_{ij}, Y_{ik} | \theta_i \stackrel{ind}{\sim} N(\mu_{xi}^* + \xi_i, \tau_{xi}^{-1}), N(\mu_{yi}^* + \xi_i, \tau_{yi}^{-1}). \quad (1)$$

We assume that the elements of θ_i are *a priori* mutually independent with priors

$\xi_i \sim N(\tilde{\mu}_\xi, \tilde{\tau}_\xi^{-1})$, $\mu_{xi}^* \sim N(\tilde{\mu}_x^*, \tilde{\tau}_x^{*-1})$, $\mu_{yi}^* \sim N(\tilde{\mu}_y^*, \tilde{\tau}_y^{*-1})$, $\tau_{xi} \sim Ga(\tilde{a}_x, \tilde{b}_x)$, and $\tau_{yi} \sim Ga(\tilde{a}_y, \tilde{b}_y)$.

Collecting terms, we denote the hyperparameter vector by

$\tilde{\theta} = (\tilde{\mu}_\xi, \tilde{\tau}_\xi, \tilde{\mu}_x^*, \tilde{\tau}_x^*, \tilde{\mu}_y^*, \tilde{\tau}_y^*, \tilde{a}_x, \tilde{b}_x, \tilde{a}_y, \tilde{b}_y)$, and the i^{th} prior by $p_1(\theta_i | \tilde{\theta})$. To complete the hierarchical model specification, we assume vague normal hyperpriors with mean 0 and

precision 0.001 for $\tilde{\mu}_x^*$, $\tilde{\mu}_y^*$, and $\tilde{\mu}_\xi$ and vague gamma hyperpriors with both shape and inverse scale 0.001 for $\tilde{\tau}_x^*$, $\tilde{\tau}_y^*$, $\tilde{\tau}_\xi^*$, $\tilde{a}_x$, $\tilde{b}_x$, $\tilde{a}_y$, and $\tilde{b}_y$, and we denote the second level prior by $p_2(\tilde{\theta} | \phi)$, where ϕ is the vector of fixed numerical parameters values that determine the hyperpriors. Denoting $\theta = (\theta_1, \dots, \theta_N)$, the posterior of θ given the p-PDGFR data $\mathbf{X}$, $\mathbf{Y}$ is proportional to the product of the patient-specific likelihoods and priors and the hyperprior,

$$p(\theta | \mathbf{X}, \mathbf{Y}) \propto \prod_{i=1}^N \left\{ \prod_{j=1}^{m_i} \mathbf{f}(\mathbf{X}_{ij} | \xi_i, \mu_{xi}^*, \tau_{xi}) \prod_{k=1}^{n_i} \mathbf{f}(\mathbf{Y}_{ik} | \xi_i, \mu_{yi}^*, \tau_{yi}) \right\} p_1(\theta_i | \tilde{\theta}) p_2(\tilde{\theta} | \phi).$$

Under this model, ξ_i accounts for correlations between pairs (X_{ij}, Y_{ik}) . Since

$E(X_{ij}) = \mu_{xi}^* + \xi_i$, each μ_{xi}^* accounts for correlations between pairs (X_{ij}, X_{ik}) , and μ_{yi}^* accounts for correlations between pairs (Y_{ij}, Y_{ik}) . These correlations may be computed analytically as functions of $\tilde{\mu}_\xi$, $\tilde{\mu}_x^*$, $\tilde{\mu}_y^*$, $\tilde{\tau}_\xi$, $\tilde{\tau}_x^*$, $\tilde{\tau}_y^*$, τ_{xi} and τ_{yi} . We will focus attention on $\delta_i = \mu_{yi} - \mu_{xi}$, the post-treatment minus pre-treatment change in within-patient mean p-PDGFR level, when used as a covariate in a time-to-event regression analysis of PFS.

2.3 Mixture model for the p-PDGFR data

To account for the bimodality observed in the within-patient p-PDGFR distributions, we now specify a more general version of the model given in section 2.2 by assuming that each within-patient pre-treatment and post-treatment p-PDGFR distribution is a mixture of two components, a left (L) and a right (R) distribution. To construct these mixture distributions, we first augment the observed p-PDGFR data $(\mathbf{X}_i, \mathbf{Y}_i)$ of the i^{th} patient with the unobserved (latent) patient-specific mixture indicators $\zeta_{xij} = 1$ if the pre-treatment value X_{ij} is drawn from the left side component distribution and 0 if from the right side. Similarly, $\zeta_{yik} = 1$ for the left and 0 for the right side of the post-treatment distribution. We assume that $\zeta_{xij} \sim \text{Bernoulli}(\lambda_{xi})$ and $\zeta_{yik} \sim \text{Bernoulli}(\lambda_{yi})$. Generalizing the non-mixture model formulation given above, we assume the following normal distributions for each of the L and R component distributions of $\mathbf{X}_i$ and $\mathbf{Y}_i$:

	Left Side		Right Side	
	Mean	Precision	Mean	Precision
Pre-treatment	$\mu_{xLi} = \mu_{xLi}^* + \zeta_i$	τ_{xLi}	$\mu_{xRi} = \mu_{xRi}^* + \zeta_i$	τ_{xRi}
Post-treatment	$\mu_{yLi} = \mu_{yLi}^* + \zeta_i$	τ_{yLi}	$\mu_{yRi} = \mu_{yRi}^* + \zeta_i$	τ_{yRi}

Under the mixture model, the patient-specific parameter vector now has the more refined structure $\theta_i = (\xi_i, \theta_{Li}, \theta_{Ri}, \lambda_i)$ where $\theta_{Li} = (\mu_{xLi}^*, \tau_{xLi}, \mu_{yLi}^*, \tau_{yLi})$, $\theta_{Ri} = (\mu_{xRi}^*, \tau_{xRi}, \mu_{yRi}^*, \tau_{yRi})$ and $\lambda_i = (\lambda_{xi}, \lambda_{yi})$. Thus, θ_i has 11 elements under the mixture model compared to 5 elements under the non-mixture model.

Given θ_i , each pair X_{ij} and Y_{ik} are again conditionally independent, but in order to specify the likelihood contribution of $(\mathbf{X}_i, \mathbf{Y}_i)$ we now also require the latent mixture indicators $\zeta_{xi} = (\zeta_{xi1}, \dots, \zeta_{xim_i})$ and $\zeta_{yi} = (\zeta_{yi1}, \dots, \zeta_{yin_i})$. Denoting the pdf of a normal distribution with mean μ and precision τ by $f_N(\cdot | \mu, \tau)$, the mixture model likelihood for the i^{th} patient's p-PDGFR data takes the form

$$\prod_{j=1}^{m_i} \left[\lambda_{xi} f_N(X_{ij} | \mu_{xLi}^* + \xi_i, \tau_{xLi}) \right]^{\zeta_{xij}} \left[(1 - \lambda_{xi}) f_N(X_{ij} | \mu_{xRi}^* + \xi_i, \tau_{xRi}) \right]^{1 - \zeta_{xij}} \\ \times \prod_{k=1}^{n_i} \left[\lambda_{yi} f_N(Y_{ik} | \mu_{yLi}^* + \xi_i, \tau_{yLi}) \right]^{\zeta_{yik}} \left[(1 - \lambda_{yi}) f_N(Y_{ik} | \mu_{yRi}^* + \xi_i, \tau_{yRi}) \right]^{1 - \zeta_{yik}}$$

To generalize the priors to account for the L and R components of the mixture, we assume that the elements of θ_i are *a priori* independent, with

$\xi_i \sim N(\tilde{\mu}_\xi, \tilde{\tau}_\xi^{-1})$, $\mu_{xLi}^* \sim N(\tilde{\mu}_{xL}^*, \tilde{\tau}_{xL}^{*-1})$, $\mu_{xRi}^* \sim N(\tilde{\mu}_{xR}^*, \tilde{\tau}_{xR}^{*-1})$, $\mu_{yLi}^* \sim N(\tilde{\mu}_{yL}^*, \tilde{\tau}_{yL}^{*-1})$, $\mu_{yRi}^* \sim N(\tilde{\mu}_{yR}^*, \tilde{\tau}_{yR}^{*-1})$, $\tau_{xLi} \sim Ga(\tilde{a}_{xL}, \tilde{b}_{xL})$, $\tau_{xRi} \sim Ga(\tilde{a}_{xR}, \tilde{b}_{xR})$, $\tau_{yLi} \sim Ga(\tilde{a}_{yL}, \tilde{b}_{yL})$, $\tau_{yRi} \sim Ga(\tilde{a}_{yR}, \tilde{b}_{yR})$, $\lambda_{xi} \sim Be(\alpha_x, \beta_x)$, and $\lambda_{yi} \sim Be(\alpha_y, \beta_y)$. We denote the hyperparameter vector by $\theta = (\theta_1, \theta_2)$ where $\tilde{\theta}_1 = (\tilde{\mu}_\xi, \tilde{\tau}_\xi, \tilde{\mu}_{xL}^*, \tilde{\tau}_{xL}^*, \tilde{\mu}_{xR}^*, \tilde{\tau}_{xR}^*, \tilde{\mu}_{yL}^*, \tilde{\tau}_{yL}^*, \tilde{\mu}_{yR}^*, \tilde{\tau}_{yR}^*)$, $\tilde{\theta}_2 = (\tilde{a}_{xL}, \tilde{b}_{xL}, \tilde{a}_{xR}, \tilde{b}_{xR}, \tilde{a}_{yL}, \tilde{b}_{yL}, \tilde{a}_{yR}, \tilde{b}_{yR}, \tilde{\alpha}_x, \tilde{\beta}_x, \alpha_y, \beta_y)$. To avoid non-identifiability in the mixture model, we require the restrictions $\mu_{xLi}^* < \mu_{xRi}^*$ and $\mu_{yLi}^* < \mu_{yRi}^*$, which formalize what can be seen clearly in Figure 1. There are now two post-treatment minus pre-treatment mean differences, $\delta_{Li} = \mu_{yLi} - \mu_{xLi}$ for the L component distributions and $\delta_{Ri} = \mu_{yRi} - \mu_{xRi}$ for the R components. We assume vague normal hyperpriors with mean 0 and precision 0.001 for $\tilde{\mu}_\xi$, $\tilde{\mu}_{xL}^*$, $\tilde{\mu}_{xR}^*$, $\tilde{\mu}_{yL}^*$ and $\tilde{\mu}_{yR}^*$, and vague gamma hyperpriors with both shape and inverse scale 0.001 for the rest of the hyperparameters.

2.4 Regression models for estimating PFS

For the regression analyses of PFS, we first fit a set of candidate time-to-event distributions to the PFS time data, including the exponential, Weibull, gamma, and log-normal. In each model, the linear term includes a main treatment effect (DI versus DP, the imatinib effect), treatment group-specific effects of Hb, Z_{2i} , and treatment group-specific effects of change in PSA, Z_{3i} . Under the non-mixture model, the linear term for patient i is

$$\eta_i^{non-mix} = \beta_0 + \beta_1 Z_{1i} + \{\beta_2 Z_{1i} + \beta_3 (1 - Z_{1i})\} Z_{2i} + \{\beta_4 Z_{1i} + \beta_5 (1 - Z_{1i})\} Z_{3i} \\ + \{\beta_6 Z_{1i} + \beta_7 (1 - Z_{1i})\} \delta_i. \quad (2)$$

The main imatinib effect is β_1 , the two treatment group-specific effects of Hb are β_2 and β_3 , the two treatment group-specific effects of change in PSA are β_4 and β_5 , and the two treatment group-specific effects of the mean change in p-PDGFR, δ_i , are β_6 and β_7 .

Under the mixture model, the linear term for patient i is

$$\eta_i^{mix} = \beta_0 + \beta_1 Z_{1i} + \{\beta_2 Z_{1i} + \beta_3 (1 - Z_{1i})\} Z_{2i} + \{\beta_4 Z_{1i} + \beta_5 (1 - Z_{1i})\} Z_{3i} \\ + \{\beta_6 Z_{1i} + \beta_7 (1 - Z_{1i})\} \delta_{Li} + \{\beta_8 Z_{1i} + \beta_9 (1 - Z_{1i})\} \delta_{Ri}. \quad (3)$$

This model includes four terms corresponding to the effects of the mean changes in each component of the mixture model for p-PDGFR, β_6 and β_7 for the effects of δ_{Li} within the two treatment arms and, similarly, β_8 and β_9 for the corresponding effects of δ_{Ri} . The formulations (2) and (3) differ from a conventional Bayesian regression model in that, for each patient, whereas each of the covariates Z_{1i} , Z_{2i} , Z_{3i} is a single value, each covariate δ_i in (2) or δ_{Li} , δ_{Ri} in (3) is itself a random parameter whose posterior distribution is estimated from the p-PDGFR data. In section 3.2, we will provide details of the model and

computational algorithm accounting for the randomness of δ_i from the non-mixture model, or $(\delta_{Li}, \delta_{Ri})$ from the mixture model.

Denoting either the non-mixture or mixture model form of the linear term by η , we considered an exponential distribution with pdf $f(t | \eta) = e^{-\eta} \exp\{-e^{-\eta}t\}$, a Weibull with pdf $f(t | \eta, v) = vt^{v-1}e^{-\eta} \exp\{-e^{-\eta}t^v\}$, a gamma with pdf $f(t | \eta, v) = \{\exp(-\eta)\}^v t^{v-1} \exp(-e^{-\eta}t)/\Gamma(v)$, and a log-normal with pdf $f(t | \eta, v) = (v/2\pi)^{1/2} t^{-1} \exp[-v\{\log(t)-\eta\}^2/2]$. We assumed vague normal prior distributions with mean 0 and precision 0.001 for the β_j 's in each linear term and a vague gamma prior distribution with both shape and inverse scale 0.001 for any of the additional scale or shape parameters v appearing in the time-to-event model pdf. To choose a model for the PFS analyses, we compared the fits of the four distributions using the Deviance Information Criterion (DIC, Spiegelhalter, Best, Carlin and van der Linde, 2002) and the Bayes Information Criterion (BIC, Schwarz, 1978).

3. Model fitting

We used Markov chain Monte Carlo (MCMC) methods to obtain samples from the posterior distributions of the parameters for both the non-mixture and mixture models (Gilks, Richardson and Spiegelhalter, 1996; McLachlan and Peel, 2000), and also for the time-to-event regression model fits. Each algorithm was run in five parallel chains to assess convergence of the MCMC algorithm. A burn-in of 1,000 and a chain of length 20,000, retaining every 10th sample, provided adequate convergence. The sampling scheme that we used to compute posteriors for the hierarchical mixture model is described in the Appendix.

3.1 Computation of the regression model parameter posterior

Since the patient-specific parameters δ_i under the non-mixture model, or δ_{iL}, δ_{iR} under the mixture model, are used as covariates in the linear terms (2) and (3) of the regression models, it is important to clarify how the posteriors are computed. For simplicity, denote the PFS time data by $\mathbf{T}$, the changes in mean p-PDGFR by δ , and the covariate parameters by β . The posterior of the regression model parameters is

$$p(\beta, v | \mathbf{T}, \mathbf{Z}, \mathbf{X}, \mathbf{Y}) = \int p(\beta, v | \mathbf{T}, \mathbf{Z}, \delta) p(\delta | \mathbf{X}, \mathbf{Y}) d\delta. \quad (4)$$

This accounts for the uncertainty in δ given the p-PDGFR data. Equation (4) relies on the assumption that δ is conditionally independent of $(\mathbf{T}, \mathbf{Z})$ given $(\mathbf{X}, \mathbf{Y})$. We computed the posterior in (4) using the method given in section 4.3 of Mwalili, Lesaffre and Declerck (2005), as follows. At each iteration of the MCMC algorithm for obtaining the posterior of (β, v) , for each patient $i = 1, \dots, N$, a value of θ_i was sampled from $p(\theta_i | \mathbf{X}, \mathbf{Y})$, and the resulting $\delta_i = \delta(\theta_i)$ under the non-mixture model or $(\delta_{Li}, \delta_{Ri}) = (\delta_{Li}(\theta_i), \delta_{Ri}(\theta_i))$ under the mixture model were incorporated into the linear component of the regression model for PFS.

3.2 Goodness-of-fit analysis for the p-PDGFR models

To compare how well the mixture and non-mixture models fit the p-PDGFR data, we used the following supnorm metric. The basic idea is to compute the maximum distance between the empirical distribution of each pre-treatment and post-treatment within-patient p-PDGFR sample and the corresponding predictive distribution under each fitted model, and then use these maximum distances to determine whether the mixture or non-mixture model provides a better fit. To do this, we first computed the empirical CDF's of $\mathbf{X}_i$ and $\mathbf{Y}_i$ for each $i = 1, \dots, N$,

$$\widehat{F}_i(x) = \frac{1}{m_i} \sum_{j=1}^{m_i} I(X_{ij} \leq x) \text{ and } \widehat{G}_i(y) = \frac{1}{n_i} \sum_{k=1}^{n_i} I(Y_{ik} \leq y),$$

where $I(A)$ denotes the indicator of the event A . Next, we fixed $x_0 < \min_{i,j}\{X_{ij}\}$ and $y_0 < \min_{i,j}\{Y_{ij}\}$ and a sufficiently small increment Δ in the domains of $\mathbf{X}$ and $\mathbf{Y}$, and computed the corresponding empirical estimated probability increments

$$Dx_{ih} \equiv \widehat{F}_i(x_0 + (h+1)\Delta) - \widehat{F}_i(x_0 + h\Delta)$$

and

$$Dy_{ih} \equiv \widehat{G}_i(y_0 + (h+1)\Delta) - \widehat{G}_i(y_0 + h\Delta)$$

for $h = 0, \dots, H$ and $i = 1, \dots, N$. We set $x_0 = y_0 = 0.49$, $\Delta = 0.02$, and $H = 75$, after taking the ranges and variability of $\mathbf{X}$ and $\mathbf{Y}$ into account. For the mixture model, given ξ_i and

$\theta_{xi} = (\mu_{xLi}^*, \mu_{xRi}^*, \tau_{xLi}^{-1}, \tau_{xRi}^{-1}, \lambda_{xi})$, the pre-treatment p-PDGFR distribution was the mixture of normals

$$f_{xi}^{mix}(x|\xi_i, \theta_{xi}) = \lambda_{xi} f_N(x|\xi_i + \mu_{xLi}^*, \tau_{xLi}^{-1}) + (1 - \lambda_{xi}) f_N(x|\xi_i + \mu_{xRi}^*, \tau_{xRi}^{-1}),$$

and we defined the estimated mixture distribution $\widehat{f}_{xi}^{mix}(x)$ to be the posterior mean of

$f_{xi}^{mix}(x|\xi_i, \theta_{xi})$, with the estimated post-treatment p-PDGFR mixture distribution $\widehat{f}_{yi}^{mix}(y)$ obtained similarly. The supnorm metrics for patient i with data $\mathbf{X}_i$ and $\mathbf{Y}_i$ are

$$Sx_i = \max_{0 \leq h \leq H} |Dx_{ih} - \widehat{f}_{xi}^{mix}(x_0 + (h+0.5)\Delta)|$$

and

$$Sy_i = \max_{0 \leq h \leq H} |Dy_{ih} - \widehat{f}_{yi}^{mix}(y_0 + (h+0.5)\Delta)|,$$

and we used $S_i = \max\{Sx_i, Sy_i\}$ for our goodness-of-fit analysis of the mixture model. For the non-mixture model, we applied the same method using the conventional uni-modal normal distributions

$$f_{xi}^{non-mix}(x|\xi_i, \mu_{xi}^*, \tau_{xi}) = f_N(x|\xi_i + \mu_{xi}^*, \tau_{xi}^{-1})$$

and

$$f_{yi}^{non-mix}(y|\xi_i + \mu_{yi}^*, \tau_{yi}) = f_N(y|\xi_i + \mu_{yi}^*, \tau_{yi}^{-1}).$$

4. Results

4.1 Fit of the non-mixture and mixture models to the p-PDGFR data

The population distributions of the patient-specific mean p-PDGFR differences between post-treatment and pre-treatment under the non-mixture and mixture models are summarized in Table 1, which gives the posterior means and standard deviations in the population-level distributions for δ_i , δ_{Li} and δ_{Ri} . The most important message in Table 1 is that the left and right component mean changes in p-PDGFR, δ_{Li} and δ_{Ri} , have very different population distributions. Comparing these two distributions with that of δ_i shows that the non-mixture model obfuscates the bimodality of the p-PDGFR distributions, and that δ_i is actually a weighted average of δ_{Li} and δ_{Ri} . An interesting property of the pre- and post-treatment mixing proportions, λ_{xi} and λ_{yi} , is that the individual posterior estimates of λ_{xi} and λ_{yi} are virtually both on average .20, but both vary substantially, roughly between .05 and .40. Moreover, the dispersion of the left component distribution is much larger than that of the right component. The posterior estimates of the patient-specific standard deviations of the left components $\tau_{xLi}^{-1/2}$ and $\tau_{yLi}^{-1/2}$ are on average about .16, while those of the right components $\tau_{xRi}^{-1/2}$ and $\tau_{yRi}^{-1/2}$ are on average about .05. These statistics reflect what can be seen in Figure 1. Thus, the hierarchical model appears to be quite appropriate for all elements of θ_i .

Computing the supnorm metrics $S_1, \dots, S_{88}$ to compare goodness-of-fit for the non-mixture and mixture models for the within-patient p-PDGFR data gave median $\{S_i\} = .098$ with 5th and 95th percentiles .048 and .155 for the non-mixture model, compared to median $\{S_i\} = .033$ with 5th and 95th percentiles .016 and .055 for the mixture model. Thus, the mixture model appears to provide a substantially better fit.

4.2 Regression analyses for PFS time

Table 2 gives the DIC and BIC values for the fitted regression models for PFS, under the mixture formulation, for each of four event time distributions. Both statistics indicate that the log-normal gives the best fit. In particular, this allows the hazard of PFS time to be non-monotone.

Table 3 summarizes three fitted log-normal regression models, first using the non-mixture model for p-PDGFR to obtain a single mean change δ_i for use as a covariate, then using the pair of mean changes δ_{Li} and δ_{Ri} obtained from the mixture model as covariates, and finally a reduced model using only the right component estimated means δ_{Ri} . In all fits, as shown earlier in the definitions of $\eta_i^{non-mix}$ and η_i^{mix} , the effects of these covariates on PFS were evaluated assuming a fully interactive model with different covariate effects within the two treatment arms. Under the Bayesian formulation, for a given covariate a coefficient β with posterior concentrated around 0 corresponds to no effect, whereas a value of $\Pr(\beta > 0 \mid data)$ close to either 0 or 1 has the interpretation that PFS changes substantively with the covariate. This corresponds to a small p-value in a frequentist analysis, but without the usual problem of interpretability. Under the log-normal model, $\beta > 0$ ($\beta < 0$) corresponds to longer (shorter) PFS, on average. Thus, the positive values of the pre-treatment Hb effects correspond to a higher pre-treatment Hb level being associated with longer PFS. The negative values of the post-treatment PSA - pre-treatment PSA effects correspond to the well-known fact that a rising PSA is associated with disease progression in prostate cancer. However, this effect was much larger in the placebo arm than in the imatinib arm, which suggests that imatinib may disrupt the effect of change in PSA on PFS time. While the mean change δ_i obtained from the non-mixture model is moderately associated with PFS, the effect of the mean change δ_{Ri} from the mixture model is quite strong in the DP arm. It thus

appears that δ_i has been deconvoluted into a strongly associated component, δ_{Ri} , and a weakly associated component, δ_{Li} , by the mixture model. This is supported by the DIC and BIC values in Table 3, which indicate that the regression model including δ_{Ri} but not δ_{Li} in each arm gives a better fit than either the regression model including both δ_{Li} and δ_{Ri} , or including only δ_i from the non-mixture model.

Figure 2 gives plots of the estimated covariate-adjusted hazard functions under the reduced log-normal model that includes δ_{Ri} but not δ_{Li} . In the figure, the left and right columns correspond to the DI and DP arms, the rows correspond to increase in PSA evaluated at its 10, 50th and 90th percentiles, and within each of the six plots the estimated hazard is given for δ_{Ri} evaluated at its 10%, mean, and 90% points. This figure may be regarded as a graphical analysis of covariance, illustrating the increasing hazard of disease progression with larger increase in PSA, a higher hazard in the imatinib arm compared to placebo, a decreasing hazard with larger δ_{Ri} , and the interactions of both the PSA increase and δ_{Ri} with treatment arm.

5. Discussion

In this analysis, we assumed patient-specific mean and precision parameters to analyze the p-PDGFR data, and introduced patient-specific mixture parameters for the purpose of deconvoluting the p-PDGFR distributions. Rather than fixing hyperparameters of priors for the patient-specific parameters, we assumed a hierarchical model. One could assume simpler, more parsimonious models, for example, a model where common data precision parameters are assumed among patients. For the p-PDGFR data, however, the goodness-of-fit analyses based on the supnorm metric indicated that the models including patient-specific precision parameters provide better fits.

The fact that the δ_{Ri} effect was much larger in the placebo arm compared to the imatinib arm, like the PSA effect, may indicate that imatinib disrupts the effect of change in p-PDGFR. It may be hypothesized that the two component distributions of phosphorylated p-PDGFR represent different subsets of leukocytes, i.e. neutrophils, monocytes or lymphocytes, with variable roles in facilitating antitumor efficacy of taxane therapy. Alternatively, they may simply represent two leukocyte subpopulations having different capacities as cellular surrogates for a pharmacodynamic signature of antitumor efficacy, but with no functional difference. A third possibility is that the left component distribution is an artifact of a problem with laboratory method, although we have not been able to identify such a source of the bimodality.

Our analyses show that, while a decline in PSA and a rise in p-PDGFR are associated with longer PFS, these effects are much smaller in the imatinib arm. These results suggest that imatinib may disrupt the effects of these covariates on PFS. To validate this apparent effect of imatinib, however, further investigation with a larger patient cohort would be required.

Acknowledgments

Satoshi Morita's work was supported in part by grant H20-CLINRES-G-009 from the Ministry of Health, Labour, and Welfare in Japan. The authors thank two referees for their constructive comments and suggestions, which led to substantive improvements in the paper.

Appendix

The MCMC sampling scheme that we used to fit the mixture model was carried out as follows. We used the Gibbs sampling from standard conditional distributions (Steps 1 to 8)

in combination with Metropolis-Hastings steps for sampling from nonstandard conditional distributions (Steps 9 and 10).

Step 0: Initialize the parameters

$\xi_i^{(0)}, \mu_{xLi}^{*(0)}, \mu_{xRi}^{*(0)}, \mu_{yLi}^{*(0)}, \mu_{yRi}^{*(0)}, \tau_{xLi}^{*(0)}, \tau_{xRi}^{*(0)}, \tau_{yLi}^{*(0)}, \tau_{yRi}^{*(0)}, \lambda_{xi}^{(0)}$ and $\lambda_{yi}^{(0)}$ for $i = 1, \dots, N$ and the hyperparameters

$\tilde{\mu}_{\xi}^{(0)}, \tilde{\mu}_{xL}^{*(0)}, \tilde{\mu}_{xR}^{*(0)}, \tilde{\mu}_{yL}^{*(0)}, \tilde{\mu}_{yR}^{*(0)}, \tilde{\tau}_{\xi}^{(0)}, \tilde{\tau}_{xL}^{*(0)}, \tilde{\tau}_{xR}^{*(0)}, \tilde{\tau}_{yL}^{*(0)}, \tilde{\tau}_{yR}^{*(0)}, \tilde{\alpha}_x^{(0)}, \tilde{\beta}_x^{(0)}, \tilde{\alpha}_y^{(0)}, \tilde{\beta}_y^{(0)}, \tilde{a}_{xL}^{(0)}, \tilde{b}_{xL}^{(0)}, \tilde{a}_{xR}^{(0)}, \tilde{b}_{xR}^{(0)}, \tilde{a}_{yL}^{(0)}, \tilde{b}_{yL}^{(0)}, \tilde{a}_{yR}^{(0)}$ and $\tilde{b}_{yR}^{(0)}$ using the within-patient pre-treatment and post-treatment empirical means and variances of the p-PDGFR samples to obtain the initial values.

For each iteration $t = 1, 2, \dots$,

Step 1: Sample the latent mixture indicator variables $\zeta_{xij}^{(t+1)}$ and $\zeta_{yik}^{(t+1)}$, for $j = 1, \dots, m_i$ and $k = 1, \dots, n_i$, respectively from Bernoulli($\zeta_{xij}^{(t)}$) and Bernoulli($\zeta_{yik}^{(t)}$) distributions, where

$$\zeta_{xij}^{(t)} = \frac{\lambda_{xi}^{(t)} f_N(X_{ij} | \mu_{xLi}^{*(t)} + \xi_i^{(t)}, \tau_{xLi}^{(t)})}{\lambda_{xi}^{(t)} f_N(X_{ij} | \mu_{xLi}^{*(t)} + \xi_i^{(t)}, \tau_{xLi}^{(t)}) + (1 - \lambda_{xi}^{(t)}) f_N(X_{ij} | \mu_{xRi}^{*(t)} + \xi_i^{(t)}, \tau_{xRi}^{(t)})},$$

$$\zeta_{yik}^{(t)} = \frac{\lambda_{yi}^{(t)} f_N(Y_{ik} | \mu_{yLi}^{*(t)} + \xi_i^{(t)}, \tau_{yLi}^{(t)})}{\lambda_{yi}^{(t)} f_N(Y_{ik} | \mu_{yLi}^{*(t)} + \xi_i^{(t)}, \tau_{yLi}^{(t)}) + (1 - \lambda_{yi}^{(t)}) f_N(Y_{ik} | \mu_{yRi}^{*(t)} + \xi_i^{(t)}, \tau_{yRi}^{(t)})},$$

and repeat this step for each $i = 1, \dots, N$.

Step 2: Sample $\xi_i^{(t+1)}$ for $i = 1, \dots, N$ from the following normal distribution,

$$N\left(\frac{M_{\xi}^{(t)}}{S_{\xi}^{(t)}}, (S_{\xi}^{(t)})^{-1}\right) \quad (5)$$

where

$$\begin{aligned}
 M_{\xi}^{(t)} &= \tilde{\tau}_{\xi}^{(t)} \tilde{\mu}_{\xi}^{(t)} \\
 &+ \tau_{xLi}^{(t)} \left(\sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)} \right) (\bar{X}_{Li} - \mu_{xLi}^{*(t)}) \\
 &+ \tau_{xRi}^{(t)} \left(\sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)}) \right) (\bar{X}_{Ri} - \mu_{xRi}^{*(t)}) + \tau_{yLi}^{(t)} \left(\sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)} \right) (\bar{Y}_{Li} - \mu_{yLi}^{*(t)}) \\
 &+ \tau_{yRi}^{(t)} \left(\sum_{k=1}^{n_i} (1 - \zeta_{yik}^{(t+1)}) \right) (\bar{Y}_{Ri} - \mu_{yRi}^{*(t)}) \text{ and } S_{\xi}^{(t)} \\
 &= \tilde{\tau}_{\xi}^{(t)} + \tau_{xLi}^{(t)} \sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)} \\
 &+ \tau_{xRi}^{(t)} \sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)}) + \tau_{yLi}^{(t)} \sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)} + \tau_{yRi}^{(t)} \sum_{k=1}^{n_i} (1 - \zeta_{yik}^{(t+1)}). \text{ And } \bar{X}_{Li} = \left(\sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)} X_{ij} \right) \\
 &/ \sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)} \\
 , \bar{X}_{Ri} &= \left(\sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)}) X_{ij} \right) \\
 &/ \left(\sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)}) \right) \\
 , \bar{Y}_{Li} &= \left(\sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)} Y_{ik} \right) \\
 &/ \sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)} \\
 , \text{ and } \bar{Y}_{Ri} \\
 &= \left(\sum_{k=1}^{n_i} (1 - \zeta_{yik}^{(t+1)}) Y_{ik} \right) \\
 &/ \left(\sum_{k=1}^{n_i} (1 - \zeta_{yik}^{(t+1)}) \right)
 \end{aligned}$$

Step 3: Sample $\mu_{xLi}^{*(t+1)}$ and $\mu_{xRi}^{*(t+1)}$ for $i = 1, \dots, N$, under the constraint $\mu_{xLi}^{*(t+1)} < \mu_{xRi}^{*(t+1)}$, using the inverse cumulative distribution method, through the following sub-steps 3-1 to 3-4.

Step 3-1: Sample $\mu_{xLi}^{(t+1)}$ from the following normal distribution,

$$N \left(\frac{M_{\mu xL}^{(t)}}{S_{\mu xL}^{(t)}}, (S_{\mu xL}^{(t)})^{-1} \right) \quad (6)$$

where $M_{\mu xL}^{(t)} = \tilde{\tau}_{xL}^{*(t)} \tilde{\mu}_{xL}^{*(t)} + \tau_{xLi}^{(t)} \left(\sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)} \right) (\tilde{X}_{Li} - \xi_i^{(t+1)})$ and $S_{\mu xL}^{(t)} = \tilde{\tau}_{xL}^{*(t)} + \tau_{xLi}^{(t)} \sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)}$.

Step 3-2: Calculate $\int_{\mu_{xLi}^{*(t+1)}}^{\infty} \phi(\mu_{xRi}^*) d\mu_{xRi}^* = 1 - \Phi(\mu_{xLi}^{*(t+1)})$, where the quantity of $\Phi(\mu_{xLi}^{*(t+1)})$ is the normal CDF at $\mu_{xLi}^{*(t+1)}$ under the following normal distribution,

$$N \left(\frac{M_{\mu xR}^{(t)}}{S_{\mu xR}^{(t)}}, (S_{\mu xR}^{(t)})^{-1} \right) \quad (7)$$

where

$$M_{\mu xR}^{(t)} = \tilde{\tau}_{xR}^{*(t)} \tilde{\mu}_{xR}^{*(t)} + \tau_{xRi}^{(t)} \left(\sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)}) \right) (\tilde{X}_{Ri} - \xi_i^{(t+1)}) \text{ and } S_{\mu xR}^{(t)} = \tilde{\tau}_{xR}^{*(t)} + \tau_{xRi}^{(t)} \sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)}).$$

Step 3-3: Sample $u_2^{(t+1)}$ from $U \left[\Phi(\mu_{xLi}^{*(t+1)}), 1 \right]$.

Step 3-4: Sample $\mu_{xRi}^{*(t+1)}$ from $\Phi^{-1}(u_2^{(t+1)})$, where $\Phi^{-1}(u_2^{(t+1)})$ is the inverse normal CDF evaluated at $u_2^{(t+1)}$ under the normal distribution (7).

Step 4: Sample $\mu_{yLi}^{*(t+1)}$ and $\mu_{yRi}^{*(t+1)}$ for $i = 1, \dots, N$, similarly to Step 3, under the restriction that $\mu_{yLi}^{*(t+1)} < \mu_{yRi}^{*(t+1)}$.

Step 5: Sample the mixture parameters $\lambda_{xi}^{(t+1)}$ and $\lambda_{yi}^{(t+1)}$ for $i = 1, \dots, N$ respectively from

$$Be \left(\tilde{\alpha}_x^{(t)} + \sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)}, \tilde{\beta}_x^{(t)} + m_i - \sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)} \right)$$

and

$$Be \left(\tilde{\alpha}_y^{(t)} + \sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)}, \tilde{\beta}_y^{(t)} + n_i - \sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)} \right),$$

where $(\tilde{\alpha}_x, \tilde{\beta}_x)$ and $(\tilde{\alpha}_y, \tilde{\beta}_y)$ denote the hyperparameters of the priors of λ_{xi} and λ_{yi} ,

respectively. The simulations in this step are subject to the constraint that $\lambda_{xi}^{(t+1)}$ and $\lambda_{yi}^{(t+1)}$ are restricted to the interval [0.001, 0.999].

Step 6: Sample $\tau_{xLi}^{(t+1)}$, $\tau_{xRi}^{(t+1)}$, $\tau_{yLi}^{(t+1)}$, and $\tau_{yRi}^{(t+1)}$ for $i = 1, \dots, N$ respectively from

$$\begin{aligned} & Ga \left(\tilde{a}_{xL}^{(t)} + \frac{\sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)}}{2}, \tilde{b}_{xL}^{(t)} + \frac{\sum_{j=1}^{m_i} \zeta_{xij}^{(t+1)} (X_{ij} - \mu_{xLi}^{*(t+1)} - \zeta_i^{(t+1)})^2}{2} \right), \\ & Ga \left(\tilde{a}_{xR}^{(t)} + \frac{\sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)})}{2}, \tilde{b}_{xR}^{(t)} + \frac{\sum_{j=1}^{m_i} (1 - \zeta_{xij}^{(t+1)}) (X_{ij} - \mu_{xRi}^{*(t+1)} - \zeta_i^{(t+1)})^2}{2} \right), \\ & Ga \left(\tilde{a}_{yL}^{(t)} + \frac{\sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)}}{2}, \tilde{b}_{yL}^{(t)} + \frac{\sum_{k=1}^{n_i} \zeta_{yik}^{(t+1)} (Y_{ik} - \mu_{yLi}^{*(t+1)} - \zeta_i^{(t+1)})^2}{2} \right), \text{ and} \\ & Ga \left(\tilde{a}_{yR}^{(t)} + \frac{\sum_{k=1}^{n_i} (1 - \zeta_{yik}^{(t+1)})}{2}, \tilde{b}_{yR}^{(t)} + \frac{\sum_{k=1}^{n_i} (1 - \zeta_{yik}^{(t+1)}) (Y_{ik} - \mu_{yRi}^{*(t+1)} - \zeta_i^{(t+1)})^2}{2} \right), \end{aligned}$$

where $(\tilde{a}_{xL}, \tilde{b}_{xL})$, $(\tilde{a}_{xR}, \tilde{b}_{xR})$, $(\tilde{a}_{yL}, \tilde{b}_{yL})$, and $(\tilde{a}_{yR}, \tilde{b}_{yR})$ denote the hyperparameters of the priors of τ_{xLi} , τ_{xRi} , τ_{yLi} , and τ_{yRi} , respectively.

Step 7: sample $\tilde{\mu}_{\xi}^{(t+1)}$, $\tilde{\mu}_{xL}^{*(t+1)}$, $\tilde{\mu}_{yL}^{*(t+1)}$, $\tilde{\mu}_{xR}^{*(t+1)}$, and $\tilde{\mu}_{yR}^{*(t+1)}$ respectively from the following normal distributions,

$$\begin{aligned}
& N\left(\frac{\tau_{\xi}^{\phi}\mu_{\xi}^{\phi}+N\tilde{\tau}_{\xi}^{(t)}\tilde{\mu}_{\xi}^{(t+1)}}{\tau_{\xi}^{\phi}+N\tilde{\tau}_{\xi}^{(t)}}, (\tau_{\xi}^{\phi}+N\tilde{\tau}_{\xi}^{(t)})^{-1}\right), \\
& N\left(\frac{\tau_{xL}^{\phi}\mu_{xL}^{\phi}+N\tilde{\tau}_{xL}^{(t)}\tilde{\mu}_{xL}^{(t+1)}}{\tau_{xL}^{\phi}+N\tilde{\tau}_{xL}^{(t)}}, (\tau_{xL}^{\phi}+N\tilde{\tau}_{xL}^{(t)})^{-1}\right), \\
& N\left(\frac{\tau_{xR}^{\phi}\mu_{xR}^{\phi}+N\tilde{\tau}_{xR}^{(t)}\tilde{\mu}_{xR}^{(t+1)}}{\tau_{xR}^{\phi}+N\tilde{\tau}_{xR}^{(t)}}, (\tau_{xR}^{\phi}+N\tilde{\tau}_{xR}^{(t)})^{-1}\right), \\
& N\left(\frac{\tau_{yL}^{\phi}\mu_{yL}^{\phi}+N\tilde{\tau}_{yL}^{(t)}\tilde{\mu}_{yL}^{(t+1)}}{\tau_{yL}^{\phi}+N\tilde{\tau}_{yL}^{(t)}}, (\tau_{yL}^{\phi}+N\tilde{\tau}_{yL}^{(t)})^{-1}\right), \text{ and} \\
& N\left(\frac{\tau_{yR}^{\phi}\mu_{yR}^{\phi}+N\tilde{\tau}_{yR}^{(t)}\tilde{\mu}_{yR}^{(t+1)}}{\tau_{yR}^{\phi}+N\tilde{\tau}_{yR}^{(t)}}, (\tau_{yR}^{\phi}+N\tilde{\tau}_{yR}^{(t)})^{-1}\right),
\end{aligned}$$

$$\begin{aligned}
\tilde{\mu}_{\xi}^{(t+1)} &= N^{-1} \sum_{i=1}^N \xi_i^{(t+1)} \\
& , \tilde{\mu}_{xL}^{*(t+1)} \\
&= N^{-1} \sum_{i=1}^N \mu_{xLi}^{*(t+1)} \\
& , \tilde{\mu}_{yL}^{*(t+1)} \\
&= N^{-1} \sum_{i=1}^N \mu_{yLi}^{*(t+1)} \\
& , \tilde{\mu}_{xR}^{*(t+1)} \\
&= N^{-1} \sum_{i=1}^N \mu_{xRi}^{*(t+1)} \text{ and } \tilde{\mu}_{yR}^{*(t+1)} \\
&= N^{-1} \sum_{i=1}^N \mu_{yRi}^{*(t+1)}.
\end{aligned}$$

where $(\mu_{\xi}^{\phi}, \tau_{\xi}^{\phi})$, $(\mu_{xL}^{\phi}, \tau_{xL}^{\phi})$, $(\mu_{xR}^{\phi}, \tau_{xR}^{\phi})$, $(\mu_{yL}^{\phi}, \tau_{yL}^{\phi})$ and $(\mu_{yR}^{\phi}, \tau_{yR}^{\phi})$ are fixed numerical mean and precision parameter values that determine the hyperpriors for $\tilde{\mu}_{\xi}$, $\tilde{\mu}_{xL}^*$, $\tilde{\mu}_{yL}^*$, $\tilde{\mu}_{xR}^*$, and $\tilde{\mu}_{yR}^*$, respectively.

Step 8: Sample $\tilde{\tau}_{\xi}^{(t+1)}$, $\tilde{\tau}_{xL}^{*(t+1)}$, $\tilde{\tau}_{yL}^{*(t+1)}$, $\tilde{\tau}_{xR}^{*(t+1)}$, and $\tilde{\tau}_{yR}^{*(t+1)}$ respectively from

$$\begin{aligned}
& Ga\left(a_{\xi}^{\phi}+N, \left\{b_{\xi}^{\phi}+\sum_{i=1}^N (\xi_i^{(t+1)}-\tilde{\mu}_{\xi}^{(t+1)})^2\right\}^{-1}\right), \\
& Ga\left(a_{xL}^{\phi}+N, \left\{b_{xL}^{\phi}+\sum_{i=1}^N (\mu_{xLi}^{*(t+1)}-\tilde{\mu}_{xL}^{*(t+1)})^2\right\}^{-1}\right), \\
& Ga\left(a_{xR}^{\phi}+N, \left\{b_{xR}^{\phi}+\sum_{i=1}^N (\mu_{xRi}^{*(t+1)}-\tilde{\mu}_{xR}^{*(t+1)})^2\right\}^{-1}\right), \\
& Ga\left(a_{yL}^{\phi}+N, \left\{b_{yL}^{\phi}+\sum_{i=1}^N (\mu_{yLi}^{*(t+1)}-\tilde{\mu}_{yL}^{*(t+1)})^2\right\}^{-1}\right), \text{ and} \\
& Ga\left(a_{yR}^{\phi}+N, \left\{b_{yR}^{\phi}+\sum_{i=1}^N (\mu_{yRi}^{*(t+1)}-\tilde{\mu}_{yR}^{*(t+1)})^2\right\}^{-1}\right).
\end{aligned}$$

The pairs of $(a_{\xi}^{\phi}, b_{\xi}^{\phi})$, $(a_{xL}^{\phi}, b_{xL}^{\phi})$, $(a_{xR}^{\phi}, b_{xR}^{\phi})$, $(a_{yL}^{\phi}, b_{yL}^{\phi})$ and $(a_{yR}^{\phi}, b_{yR}^{\phi})$ are fixed numerical shape and inverse scale parameter values that determine the hyperpriors for $\tilde{\tau}_{\xi}$, $\tilde{\tau}_{xL}^*$, $\tilde{\tau}_{yL}^*$, $\tilde{\tau}_{xR}^*$ and $\tilde{\tau}_{yR}^*$, respectively.

Step 9: Generate $(\tilde{\alpha}_x^{(t+1)}, \tilde{\beta}_x^{(t+1)})$ and $(\tilde{\alpha}_y^{(t+1)}, \tilde{\beta}_y^{(t+1)})$ using the the Metropolis-Hastings algorithm. The the Metropolis-Hastings algorithm was implemented using a product of two independent log-normal distributions as a proposal distribution.

Step 9-1: Draw a sample as a candidate for new $\tilde{\alpha}_x$ from a log-normal distribution with mean $m_{\tilde{\alpha}_x}^{(t)}$ and standard deviation $\sigma_{\tilde{\alpha}_x}$ fixed at 0.1 on the log scale, $m_{\tilde{\alpha}_x}^{(t)}$ is given by $\log(\tilde{\alpha}_x^{(t)}) + \sigma_{\tilde{\alpha}_x}^2$. Perform the same sampling for $\tilde{\beta}_x$ with $\sigma_{\tilde{\beta}_x}$ fixed at 0.15.

Step 9-2: Compute a ratio $r^{(t+1)}$ denned as

$$r^{(t+1)} = \frac{p(\tilde{\alpha}_x^{new}, \tilde{\beta}_x^{new} | \lambda_x) q(\tilde{\alpha}_x^{(t)}, \tilde{\beta}_x^{(t)} | \tilde{\alpha}_x^{new}, \tilde{\beta}_x^{new})}{p(\tilde{\alpha}_x^{(t)}, \tilde{\beta}_x^{(t)} | \lambda_x) q(\tilde{\alpha}_x^{new}, \tilde{\beta}_x^{new} | \tilde{\alpha}_x^{(t)}, \tilde{\beta}_x^{(t)})}$$

where $q(\cdot)$ denotes a proposal distribution and $\lambda_x = (\lambda_{x1}, \dots, \lambda_{xN})$.

Step 9-3: Accept the new candidates with a probability $\min(1, r^{(t+1)})$, otherwise the values of $\tilde{\alpha}_x$ and $\tilde{\beta}_x$ remain unchanged.

Step 9-4: Generate $\tilde{\alpha}_y^{(t+1)}$ and $\tilde{\beta}_y^{(t+1)}$, similarly to Steps 9-1, 9-2, 9-3.

Step 10: Generate $(\tilde{a}_{xL}^{(t+1)}, \tilde{b}_{xL}^{(t+1)})$, $(\tilde{a}_{xR}^{(t+1)}, \tilde{b}_{xR}^{(t+1)})$, $(\tilde{a}_{yL}^{(t+1)}, \tilde{b}_{yL}^{(t+1)})$, and $(\tilde{a}_{yR}^{(t+1)}, \tilde{b}_{yR}^{(t+1)})$ using the Metropolis-Hastings algorithms, similarly to Step 9.

References

- Belin TR, Rubin DB. The analysis of repeated-measures data on schizophrenic reaction times using mixture models. *Statistics in Medicine* 1995;14:747–768. [PubMed: 7644856]
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian Data Analysis*. 2nd Edition. New York: Chapman and Hall/CRC; 2004.
- Gilks, W.; Richardson, S.; Spiegelhalter, D. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall; 1996.
- Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970;57:97–109.
- Ibrahim, JG.; Chen, M-H.; Sinha, D. *Bayesian Survival Analysis*. New York: Springer Verlag; 2001.
- Lindsey, JK. *Applying Generalized Linear Models*. New York: Springer; 2000.
- Mathew P, Thall PF, Bucana CD, Oh WK, Morris MJ, Jones DM, et al. Platelet-derived growth factor receptor inhibition and chemotherapy for castration-resistant prostate cancer with bone metastases. *Clinical Cancer Research* 2007;13:5816–5824. [PubMed: 17908974]
- McLachlan, G.; Peel, D. *Finite Mixture Models*. New York: Wiley; 2000.
- Mwalili SM, Lesaffre E, Declerck D. A Bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study. *J. R. Statist. Soc. C* 2005;54:77–93.
- Pietras K, Sjoblom T, Rubin K, Heldin C-H, Ostman A. PDGF receptors as cancer drug targets. *Cancer Cell* 2003;3:439–443. [PubMed: 12781361]
- Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978;6:461–464.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 2002;64:583–639.
- Uehara H, Kim SJ, Karashima T, David LS, Dominic F, Tsan R, et al. Effects of blocking platelet-derived growth factor-receptor signaling in a mouse model of experimental prostate cancer bone metastases. *Journal of the National Cancer Institute* 2003;95:458–570. [PubMed: 12644539]

Figure 1.
Histograms of the log-transformed p-PDGFR values for each of three typical patients,

Figure 2.

Plots of the estimated hazard functions under the final log-normal mixture model, given in Table 3, that includes δ_{Ri} but not δ_{Li} . Within each of the six plots, the three hazards correspond to δ_{Ri} set equal to its 10% (dotted line), mean (solid line), and 90% (dashed line) points. The remaining covariate, Hb, is set equal to its mean in all plots.

Table 1

Summary of the posterior distributions of the means and standard deviations (SDs) of the population-level parameters for the patient-specific post-treatment minus pre-treatment mean p-PDGFR differences, obtained under either the non-mixture model or the mixture model.

Model	Population-level Parameter	Posterior Quantities		
		Mean	5%	95%
Non-mixture	$\text{Mean}(\delta_i) = \mu_y^* - \mu_x^*$	0.016	0.008	0.024
	$\text{SD}(\delta_i) = (\tau_y^{*-1} + \tau_x^{*-1})^{1/2}$	0.047	0.042	0.053
Mixture	$\text{Mean}(\delta_{Li}) = \mu_{yL}^* - \mu_{xL}^*$	0.035	0.032	0.038
	$\text{SD}(\delta_{Li}) = (\tau_{yL}^{*-1} + \tau_{xL}^{*-1})^{1/2}$	0.070	0.060	0.087
	$\text{Mean}(\delta_{Ri}) = \mu_{yR}^* - \mu_{xR}^*$	0.007	0.006	0.009
	$\text{SD}(\delta_{Ri}) = (\tau_{yR}^{*-1} + \tau_{xR}^{*-1})^{1/2}$	0.053	0.047	0.058

Table 2

Deviance information criterion (DIC) and Bayesian information criterion (BIC) for four competing event time distributions used to estimate progression free survival time under the mixture model.

Distribution	DIC	BIC
Exponential	417.62	451.10
Weibull	405.67	442.38
Gamma	405.16	442.37
Log-normal	397.03	433.19

Table 3

Fitted log-normal non-mixture and mixture regression models for estimating progression-free survival time. The reduced mixture model includes the right-hand mean changes but not the left-hand mean changes in p-PDGFR. The posterior mean of each regression coefficient is denoted by $\hat{\beta}$, with its standard deviation (SD) given in parentheses. DIC = deviance information criterion, BIC = Bayesian information criterion, DI = docetaxel + imatinib arm, and DP = docetaxel + placebo arm.

Variable	Non-mixture Model		Mixture Model		Reduced Mixture Model	
	$\hat{\beta}$ (SD)	$\Pr(\hat{\beta} > 0 \text{data})$	$\hat{\beta}$ (SD)	$\Pr(\hat{\beta} > 0 \text{data})$	$\hat{\beta}$ (SD)	$\Pr(\hat{\beta} > 0 \text{data})$
Intercept	1.02 (0.37)	0.997	1.08 (0.38)	0.997	1.11 (0.36)	0.999
DI vs DP	-0.04 (0.60)	0.468	-0.15 (0.62)	0.405	-0.16 (0.59)	0.388
Hb in DI	0.87 (0.51)	0.956	0.93 (0.51)	0.968	0.91 (0.50)	0.964
Hb in DP	0.76 (0.40)	0.970	0.68 (0.41)	0.952	0.68 (0.39)	0.960
PSA in DI	-0.27 (0.18)	0.059	-0.24 (0.18)	0.083	-0.24 (0.18)	0.080
PSA in DP	-0.33 (0.14)	0.010	-0.33 (0.15)	0.010	-0.35 (0.14)	0.005
δ_i in DI	1.61 (2.64)	0.732	—	—	—	—
δ_i in DP	3.26 (3.63)	0.819	—	—	—	—
δ_{Li} in DI	—	—	0.06 (1.30)	0.519	—	—
δ_{Li} in DP	—	—	0.84 (1.18)	0.766	—	—
δ_{Ri} in DI	—	—	3.29 (3.48)	0.830	3.30 (3.40)	0.836
δ_{Ri} in DP	—	—	5.72 (3.57)	0.947	5.29 (3.52)	0.936
DIC	395.15		397.03			393.07
BIC	425.07		433.19			422.70