Bayesian Personalized Treatment Selection for Advanced Breast Cancer

Peter F. Thall
Department of Biostatistics
Univ of Texas MD Anderson Cancer Center
Houston, Texas, USA

Bayesian Biostatistics Conference
Bayes 2024
Rockville, Maryland
23-25 October 2024

Collaborators

Juhee Lee, PhD, Department of Statistics, UC Santa Cruz

Pavlos Msaouel, MD, PhD, Departments of Genitourinary Oncology and Translational Molecular Pathology, MD Anderson Cancer Center

Bora Lim, MD, Department of Breast Medical Oncology, MD Anderson Cancer Center

A Common Limitation of the Medical Literature

Typical medical papers on randomized clinical trials include estimated effects of treatments and covariates on outcomes.

This falls short of what practicing physicians need to make informed treatment decisions.

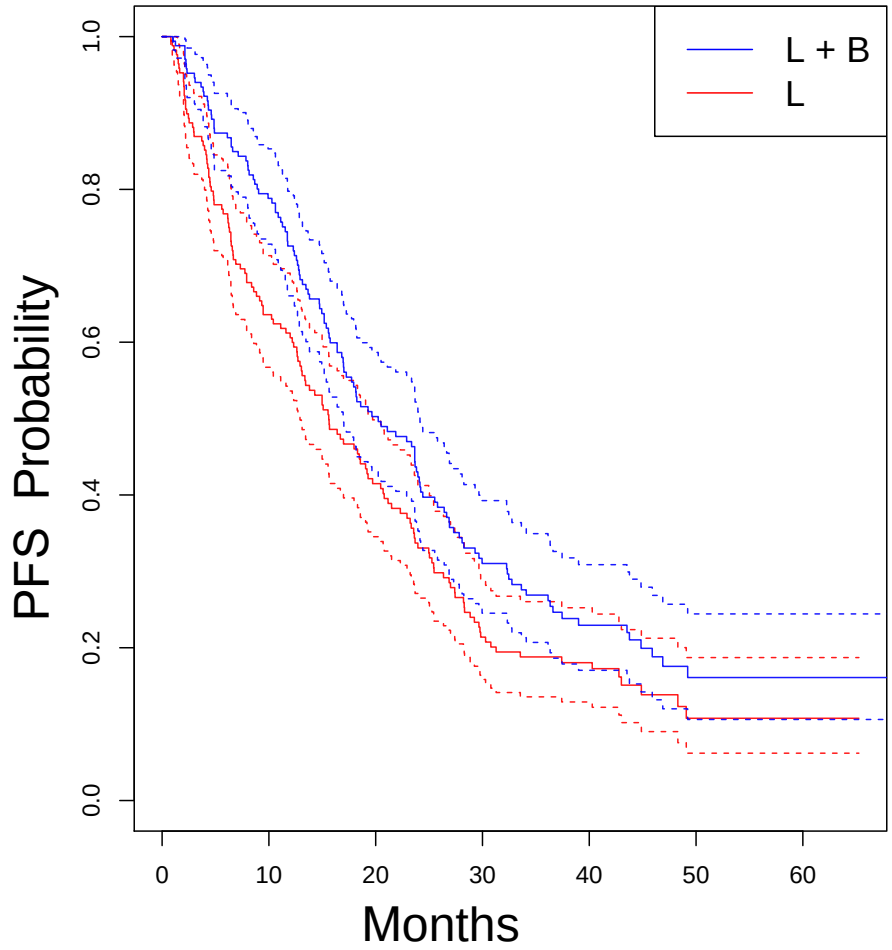
A Phase III Breast Cancer Trial: 340 patients with hormone receptor positive advanced breast cancer were randomized to
Letrozole + Bevacizumab (L+B) or
Letrozole + Placebo (L) (Dickler, et al. *JCO*, 2016).

Primary Endpoint: Progression-free survival (PFS) time

Median PFS was **20.2 mos** (95% CI 17.0 – 24.1) with L+B
versus **15.6 mos** (95% CI 12.9 – 19.7) with L.

Secondary Endpoints: 21 different types of toxicity

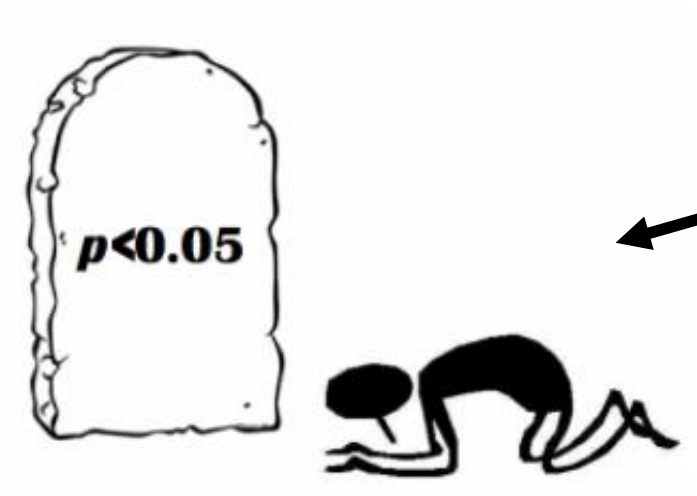
46.8% of L+B pats had ≥ 1 severe toxicity
versus **14.2%** with L



Kaplan-Meier estimates of progression free survival time distributions for $L + B$ and L in the breast cancer trial dataset.

The one-sided p-value for a test comparing PFS between L+B and L was 0.016.

Since 0.016 was smaller than the sacred value 0.05,
Dickler et al. (2016) concluded that
L+B provided a “statistically significant” improvement over L in PFS



Modern
Medical Science

A Decision Problem for Practicing Oncologists

▶ In terms of **PFS time**, $L + B$ was *slightly better* than L (Median PFS **20.2** months versus **15.6** months).

▶ In terms of **Severe Toxicity**, L was *much better* than $L + B$ (**14.2%** versus **46.8%**).

Key Baseline Covariates

$X_1 =$ Age
$X_2 =$ I[measurable disease at enrollment]
$X_3 =$ I[disease free interval prior to trial entry $\geq$ 24 months]

Question 1: How can a physician and patient account for PFS time and 21 types of toxicity when choosing a treatment?

Question 2: How can a patient's baseline covariates be used to make a *personalized* treatment choice?

Preliminaries on Utility Functions

Examples of Co-Primary Outcome Vectors

1. (**PFS Time,** $Z_1, \cdots, Z_{21}$) in the breast cancer trial
2. (**Response, Toxicity**) for a phase I-II dose-finding trial
3. (**Post Operative Morbidity, Days in Hospital**) following surgery in a nutritional prehabilitation trial
4. The times to (**Toxicity, Cytokine Storm, Response, Disease Progression, Death**) in a cell therapy trial

Establishing A Utility Function

For all possible values $\boldsymbol{y}$ of $\boldsymbol{Y}$, **elicit $U(\boldsymbol{y})$ = The desirability of $\boldsymbol{y}$**

For τ = treatment and probability distribution $f_{\boldsymbol{Y}}(\boldsymbol{y}|\tau, \theta)$ with parameters θ, the Mean Utility of τ is

$$\overline{U}(\tau, \theta) \;=\; E_{\boldsymbol{Y}}\{U(\boldsymbol{Y}) \mid \tau, \theta\} \;=\; \int_{\boldsymbol{y}} U(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}|\tau, \theta) d\boldsymbol{y}.$$

Estimating Utilities

Bayesian Estimator of the Mean Utility of Treatment τ

Under a Bayesian model, given observed data $\mathcal{D}$, the Posterior Predictive Mean Utility of τ for a future patient is

$$u(\tau, \mathcal{D}) = E_\theta\{\overline{U}(\tau, \theta) \mid \mathcal{D}\} = \int_\theta \overline{U}(\tau, \theta) p(\theta \mid \mathcal{D}) d\theta.$$

Frequentist Estimator of the Mean Utility of Treatment τ

Compute a consistent estimator $\widehat{\theta}^{freq} = \widehat{\theta}(\mathcal{D})$ and plug it into the mean utility function:

$$\widehat{u}^{freq}(\tau, \mathcal{D}) = \overline{U}(\tau, \widehat{\theta}^{freq})$$

Toy Example of a Utility in the 2×2 Case

For binary outcomes $Y_R = \mathrm{I}[\text{Response}]$, $Y_T = \mathrm{I}[\text{Toxicity}]$, the four possible values of $\boldsymbol{Y} = (Y_R, Y_T)$ are (1,0), (1,1), (0,0), or (0,1).

Establishing U

- Set $U(1,0) = 100$ for the best possible outcome and $U(0,1) = 0$ for the worst possible outcome.
- Elicit the intermediate values $U(0,0)$ and $U(1,1)$.

$U(1,1) = 70$ and $U(0,0) = 40 \Longrightarrow$

	No Toxicity	Toxicity
No Response	40	0
Response	100	70

Computing Mean Utilities in the Toy Example

For $\pi_{a,b}(\tau, \theta) = \Pr(Y_R = a, Y_T = b \mid \tau, \theta)$, $a, b \in \{0, 1\}$, the **Mean Utility of Treatment** τ is

$$\overline{U}(\tau, \theta) = \sum_{a=0}^{1} \sum_{b=0}^{1} U(a, b) \pi_{a,b}(\tau, \theta)$$

$\boldsymbol{\pi}(\tau_1) = (\pi_{0,0}(\tau_1), \pi_{0,1}(\tau_1), \pi_{1,0}(\tau_1), \pi_{1,1}(\tau_1)) = (.40, .10, .30, .20)$
$\Longrightarrow \qquad \overline{U}(\tau_1, \theta) = \mathbf{60}$

$\boldsymbol{\pi}(\tau_2) = (.60, .10, .10, .20) \Longrightarrow \qquad \overline{U}(\tau_2, \theta) = \mathbf{48}.$

- $U(\boldsymbol{y})$ is subjective, which is highly desirable

- When analyzing data, one may consider different utilities and assess sensitivity of posterior inferences to $U_1(\boldsymbol{y})$ versus $U_2(\boldsymbol{y})$

Incorporating Patient Covariates

The Pavlos Msaouel Elaboration: Elicit "precision" utilities that vary with both $\boldsymbol{Y}$ and covariates $\boldsymbol{X} = (X_1, \cdots, X_p)$ to define a

Precision Utility Function Family (PUFF) $\{U(\boldsymbol{Y}, \boldsymbol{X}) : \boldsymbol{X} \in \mathcal{X}\}$

The mean utility of treatment τ for a patient with covariates $\boldsymbol{X}$ is

$$\overline{U}(\tau, \boldsymbol{X}, \theta) = \int_{\boldsymbol{y}} U(\boldsymbol{y}, \boldsymbol{X}) f_{\boldsymbol{Y}}(\boldsymbol{y} | \tau, \boldsymbol{X}, \theta) d\boldsymbol{y}.$$

Averaging over the posterior of θ gives a statistical criterion for "precision medicine": Choose a treatment τ for given $\boldsymbol{X}$ based on

$$u(\tau, \boldsymbol{X}, \mathcal{D}) = \int_{\theta} \overline{U}(\tau, \boldsymbol{X}, \theta) p(\theta \mid \mathcal{D}) d\theta.$$

Practical Advantages of a Utility Function

1. **Quantify Trade-Offs**: $U(y_R, y_T)$ = Desirability of a possible combination of y_R = response and y_T = toxicity.

2. **Reduce Dimension to Facilitate Making Decisions**: Multi-dimensional $\boldsymbol{Y} \longrightarrow$ 1-dimensional $U(\boldsymbol{Y})$, which is used to compute 1-dimensional statistics $u(\tau_1, \mathcal{D}), \cdots, u(\tau_J, \mathcal{D})$ for evaluating and comparing treatments $\tau_1, \cdots, \tau_J$.

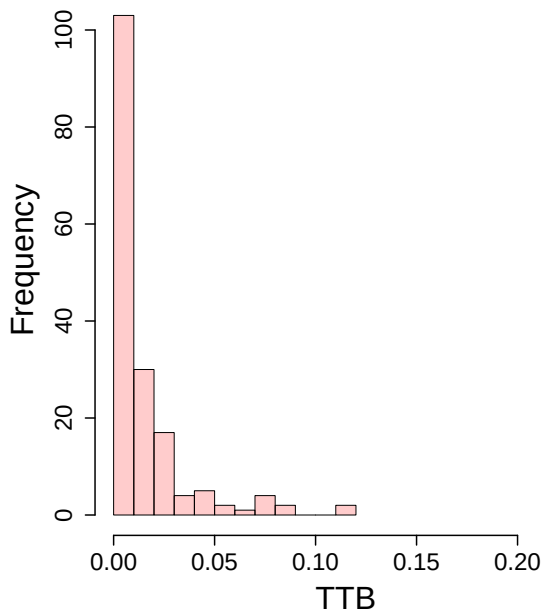
$$\text{If } u(\tau_1, \mathcal{D}) > u(\tau_2, \mathcal{D}) \text{ then } \tau_1 \text{ is more desirable than } \tau_2$$

3. **A PUFF Accounts for Heterogeneity**: $u(\tau, \boldsymbol{X}, \mathcal{D})$ is a criterion for choosing a best treatment for a patient with covariates $\boldsymbol{X}$.

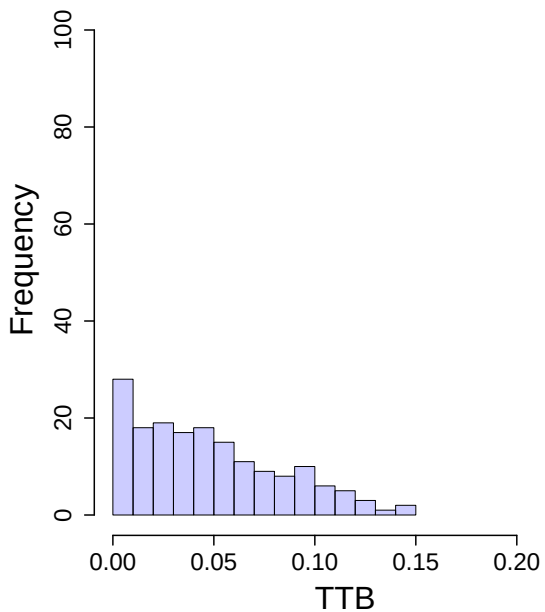
Total Toxicity Burden

To summarize 21 different types of toxicity $\boldsymbol{Z} = (Z_1, \cdots, Z_{21})$, where each Z_j = severity grade 0 (None), 1, 2, 3, 4, or 5 (Fatal), the **Total Toxicity Burden** is

$$\text{TTB} = \frac{1}{21 \times 5} \sum_{k=1}^{21} Z_k$$



TTB for L



TTB for $L + B$

Notation for Outcomes and Treatments

For each patient $i = 1, \cdots, n = 340$ in the breast cancer trial, the outcome data consisted of

T_i = PFS time, $T_i^o = T_i$ or administrative censoring time

$\varepsilon_i = \mathrm{I}[T_i^o = T_i]$.

$\tilde{T}_i = \log(T_i)$ was used in the analysis

Toxicity grades $\boldsymbol{Z}_i = (Z_{i,1}, \cdots, Z_{i,21})$ were summarized by

$Q_i = \mathrm{TTB}(\boldsymbol{Z}_i) \Longrightarrow \boldsymbol{Y}_i = (\tilde{T}_i, Q_i) = (\log \mathrm{PFS}_i, \mathrm{TTB}_i)$

$\tau_i = 1$ if patient i treated with $L + B$, $\tau_i = 0$ if treated with L

Bayesian Nonparametric Models: A Brief Review

Ferguson (1973) proposed putting a prior on a probability distribution G, rather than the usual approach of assuming a parametric model for G and putting a prior on its parameters.

He defined a Dirichlet process (DP) prior on G, characterized by a base probability measure $G_0 = E(G)$ and total mass parameter $\alpha > 0$ so that, for any measurable partition $\{B_1, \cdots, B_k\}$ of the domain of G,

$$G(B_1), \cdots, G(B_k) \sim Dirichelt(\alpha G_0(B_1), \cdots, \alpha G_0(B_k))$$

This prior is denoted $G \sim DP(G_0, \alpha)$.

Conjugacy: Given a sample $Y_1, \cdots, Y_n \overset{iid}{\sim} G$, denoting the point mass on Y by δ_Y, if $G \sim DP(G_0, \alpha)$ then the posterior is

$$G \mid Y_1, \cdots, Y_n \sim DP\left(\frac{\alpha}{\alpha + n}\, G_0 + \sum_{j=1}^{n} \frac{1}{\alpha + n}\, \delta_{Y_j},\ \alpha + n\right)$$

Stick-Breaking Construction of a DP

Sethuraman (1994) constructed a DP by using probabilities $v_1, v_2, \ldots \overset{iid}{\sim} beta(1, \alpha)$ to define weights $w_k = v_k \prod_{r<k}(1 - v_r)$
$\Longrightarrow w_1 = v_1$, $w_2 = v_2(1 - v_1)$, $w_3 = v_3(1 - v_2)(1 - v_1)$, and so on.

Given $\{w_1, w_2, \cdots\}$ and $Y_1, Y_2, \cdots \overset{iid}{\sim} G_0$, Sethuraman proved that a DP is a weighted average of point masses $\{\delta_{Y_1}, \delta_{Y_2}, ...\}$

$$G = \sum_{k=1}^{\infty} w_k \delta_{Y_k} \quad \sim \quad DP(G_0, \alpha).$$

This constructive definition greatly facilitates computing.

The "stick" is a metaphor for a probability stick of length 1, with pieces of diminishing sizes $w_1, w_2, \cdots$ successively broken off.

Dirichlet Process Mixtures

Problem: A DP prior $G \sim DP(G_0, \alpha)$ has discrete support.

Solution: To obtain continuous support, a DP mixture (**Lo**, 1984) is defined by using a continuous density function $f(y \mid \theta)$ as a smoothing kernel :

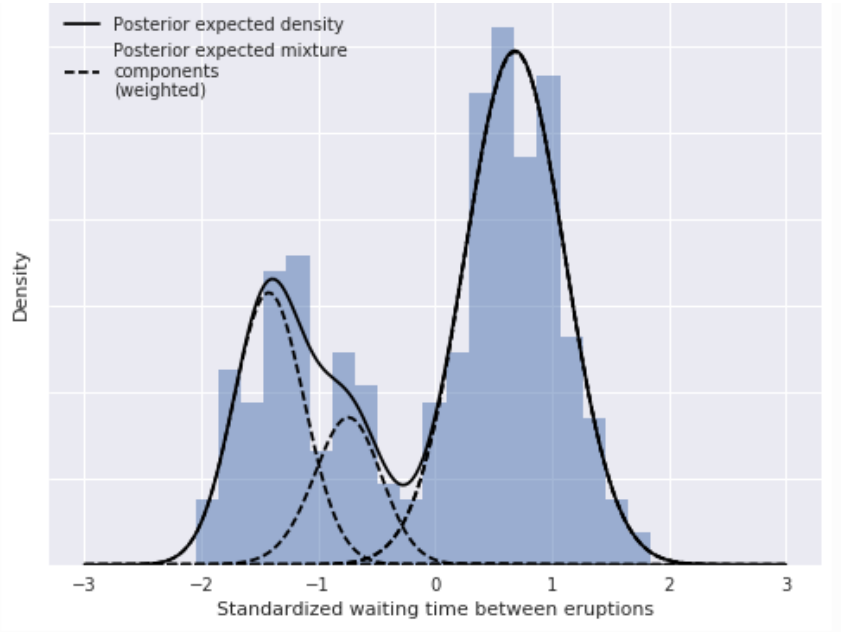
$$Y \mid G, f \sim \int f(y \mid \theta) G(d\theta).$$

A popular kernel choice is a Gaussian pdf $f(y \mid \theta) = \phi(y \mid \mu, \sigma^2)$:

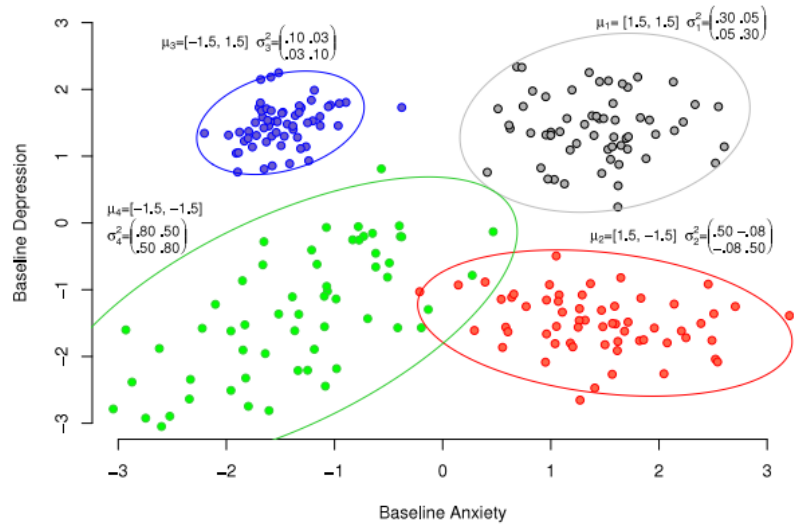
$$Y \mid G, \phi \sim \int \phi(y \mid \mu, \sigma^2) G(d\mu)$$

possibly with additional priors on μ and σ^2

Examples of DP Mixture Density Estimation



Time Between Volcanic Eruptions



Anxiety and Depression

Dependent Dirichlet Processes

To incorporate covariates $\boldsymbol{X} = (X_1, \cdots, X_p)$ into a DP **McEachern** (1999) defined a Dependent Dirichlet process (DDP) as **a fully nonparametric regression model**

$$Y_i \mid \boldsymbol{X}_i \overset{ind}{\sim} G_{\boldsymbol{X}_i}, \quad i = 1, \cdots, n,$$

where $G_{\boldsymbol{X}}$ is a DP that can vary in any way with $\boldsymbol{X}$.

A DDP is constructed by indexing everything with $\boldsymbol{X}$ in the stick-breaking algorithm: $\boxed{v_m(\boldsymbol{X}) \overset{ind}{\sim} beta(1, \alpha_{\boldsymbol{X}})}$ and $\boxed{w_m(\boldsymbol{X}) = v_m(\boldsymbol{X}) \prod_{r<m}(1 - v_r(\boldsymbol{X}))} \Longrightarrow$

$$G_{\boldsymbol{X}} = \sum_{m=1}^{\infty} w_m(\boldsymbol{X}) \delta_{\eta_m(\boldsymbol{X})}$$

A linear DDP has simple parametric linear combinations $\boxed{\eta_m(\boldsymbol{X}) = \sum_{j=1}^{p} \theta_{m,j} X_j}$ with a prior on $\boldsymbol{\theta}_m$ for each m

Application to the Breast Cancer Data

Patient Frailties: Account for heterogeneity between patients not explained by covariates $\{\boldsymbol{X}_i\}$ by defining 22-dimensional real-valued frailties $\boldsymbol{s}_i = (s_{i,0}, s_{i,1}, \ldots, s_{i,21})$ for patients $i = 1, \cdots, n$:

$$\boldsymbol{s}_i \mid \Omega \overset{iid}{\sim} \mathrm{N}_{22}(\mathbf{0}, \Omega) \quad \text{and} \quad \Omega \sim \text{Inv-Wishart}(a_s, \Omega^0).$$

Multivariate Probit Model For Ordinal Toxicity Grades

Define 21−variate latent normal frailty $\tilde{\boldsymbol{Z}}_i = (\tilde{Z}_{i,1}, \cdots, \tilde{Z}_{i,21})$. For the k^{th} toxicity, specify cut-offs $u_{k,0} < u_{k,1} < \cdots < u_{k,J}$ and define

$$Z_{i,k} = j \quad \text{if and only if} \quad u_{k,j} < \tilde{Z}_{i,k} \leq u_{k,j+1}.$$

The joint distribution of $\tilde{\boldsymbol{Z}}_i$ induces a joint distribution on $\boldsymbol{Z}_i$

Denote the 1+21=22 variate pdf of $\tilde{T}_i = \log(T_i)$ and $\tilde{\boldsymbol{Z}}_i$ by

$$h(\tilde{T}_i, \tilde{\boldsymbol{Z}}_i \mid \tau_i, \boldsymbol{X}_i, \boldsymbol{s}_i)$$

To analyze the breast cancer data, a linear DDP with covariates $(\tau, \boldsymbol{X})$ was defined, with linear terms

$$\eta_0(\tau_i, \boldsymbol{X}_i) = \beta_0 \tau_i + \sum_{r=1}^{3} \beta_r X_{i,r} \quad \text{for } \tilde{T} = \log(T) = \log(PFS)$$

$$\eta_k(\tau_i, \boldsymbol{X}_i) = \alpha_{k,0} \tau_i + \sum_{r=1}^{3} \alpha_{k,r} X_{i,r}, \quad \text{for toxicities } Z_1, \cdots, Z_{21}$$

Denote $\boldsymbol{\eta} = (\eta_0, \eta_1, \cdots, \eta_{21})$ and 22-variate normal pdf ϕ_{22}.

The linear DDP mixture model for the joint pdf h is

$$h(\tilde{T}_i, \tilde{\boldsymbol{Z}}_i \mid \tau_i, \boldsymbol{X}_i, \boldsymbol{s}_i) = \int \phi_{22}(\tilde{T}_i, \tilde{\boldsymbol{Z}}_i \mid \boldsymbol{\eta}(\tau_i, \boldsymbol{X}_i) + \boldsymbol{s}_i, \Sigma) G_{\tau, \boldsymbol{X}}(d\boldsymbol{\eta}).$$

The DDP prior on the vector $\boldsymbol{\eta} = (\eta_0, \eta_1, \cdots, \eta_{21})$ is

$$G_{\tau, \boldsymbol{X}} = \sum_{m=1}^{\infty} w_m \delta_{\boldsymbol{\eta}_m(\tau, \boldsymbol{X})}$$

Under the DDP, the joint 22-variate pdf of $\tilde{T}$ and latent toxicity grade generators $\tilde{\boldsymbol{Z}}_i = (\tilde{Z}_{i,1}, \cdots, \tilde{Z}_{i,21})$ is a weighted average of 22-variate normal linear regression model pdfs:

$$h(\tilde{T}_i, \tilde{\boldsymbol{Z}}_i \mid \tau_i, \boldsymbol{X}_i, \boldsymbol{s}_i) = \sum_{m=1}^{\infty} w_m \, \phi_{22}(\tilde{T}_i, \tilde{\boldsymbol{Z}}_i \mid \boldsymbol{\eta}_m(\tau_i, \boldsymbol{X}_i) + \boldsymbol{s}_i, \Sigma),$$

For the covariate and treatment parameters, assume priors

$$\boldsymbol{\beta}_m \overset{iid}{\sim} \mathrm{N}_{p+2}(\bar{\boldsymbol{\beta}}, \tau^2 \mathrm{I}_{p+2}), \quad \boldsymbol{\alpha}_{m,k} \overset{iid}{\sim} \mathrm{N}_{p+2}(\bar{\boldsymbol{\alpha}}_k, V), \; k = 1, \ldots, 21,$$

where $\bar{\boldsymbol{\beta}}$, $\bar{\boldsymbol{\alpha}}_k$, τ^2 and V are fixed hyperparameters.

h has parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_K)$, and patient frailties $\{\boldsymbol{s}_i\}$

Joint Posterior of $\boldsymbol{\theta}$ and patient frailties $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n$

Recall that $\varepsilon_i = \mathrm{I}[T_i^o = T_i]$ and $\tilde{T}_i = \log(T_i)$.

The **Joint Likelihood** for the i^{th} patient's observed outcomes is

$$\mathcal{L}(\tilde{T}_i^o, \varepsilon_i, \boldsymbol{Z}_i \mid \tau_i, \boldsymbol{X}_i, \boldsymbol{s}_i, \boldsymbol{\theta})$$

$$= \left\{ f_T(\tilde{T}_i^o \mid \tau_i, \boldsymbol{X}_i, s_{i,0}, \boldsymbol{\theta}) \right\}^{\varepsilon_i} \left\{ 1 - F_T(\tilde{T}_i^o \mid \tau_i, \boldsymbol{X}_i, s_{i,0}, \boldsymbol{\theta}) \right\}^{1-\varepsilon_i}$$

$$\times \prod_{k=1}^{21} p(Z_{ik} \mid \tau_i, \boldsymbol{X}_i, s_{i,k}, \boldsymbol{\theta})$$

$\Longrightarrow$ **The Joint Posterior of $\boldsymbol{\theta}$ and $\boldsymbol{s} = (\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n)$ is**

$$p(\boldsymbol{\theta}, \boldsymbol{s} \mid \mathcal{D}_n) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} \mathcal{L}(\tilde{T}_i^o, \varepsilon_i, \boldsymbol{Z}_i \mid \tau_i, \boldsymbol{X}_i, \boldsymbol{s}_i, \boldsymbol{\theta}) \times p(\boldsymbol{s}_i \mid \boldsymbol{\theta})$$

How the DDP Regression Model Works

1. The 21-variate normal distribution of *latent* $\tilde{\boldsymbol{Z}}_i$ induces a 21-variate probit distribution on *observed* toxicity grades $\boldsymbol{Z}_i$.

2. The multivariate normal priors for $\boldsymbol{\alpha}_{m,1}, \cdots, \boldsymbol{\alpha}_{m,21}$ borrow information across the toxicities.

3. The model incorporates treatment τ and covariates $\boldsymbol{X}$ linearly in the mean of each normal summand in the linear DDP.

4. The joint DDP density $h(\tilde{T}, \tilde{\boldsymbol{Z}} \mid \tau, \boldsymbol{X}, \boldsymbol{\theta})$ is a weighted average of 22-variate normals $\Longrightarrow$ The model accommodates complex interactions among τ and the entries of $\boldsymbol{X}$, and multiple modes. It is far more robust than a Cox model or accelerated failure time model for $\tilde{T}$.

Constructing a Utility Function

A Precision Utility Function Family (PUFF) indexed by $\boldsymbol{X}$ is :

$$U_{Total}(T, Q, \boldsymbol{X}) = U_{PFS}(T) \times U_{TTB}(Q, \boldsymbol{X})$$

where $U_{PFS}(T)$ increases with T and $0 < U_{TTB}(Q, \boldsymbol{X}) < 1$

$U_{TTB}(Q, \boldsymbol{X})$ is a multiplicative penalty for Q = TTB

$$0 < U_{PFS}(T) < 100$$

No TTB penalty if $Q = 0$ (No toxicity): $U_{TTB}(0, Age) = 1$ and $U_{Total}(T, Q, Age) = U_{PFS}(T)$ regardless of Age

Rationale for the Utility

1. Only $X_1 =$ Age is included in U_{TTB}, because

- Younger advanced breast cancer patients care more about extending PFS than controlling TTB

- Older patients have shorter expected survival time, so they care more about maintaining a good quality of life: A sightly shorter PFS time is an acceptable tradeoff for a much lower TTB $\Longrightarrow$

$$U_{Total}(T, Q, \boldsymbol{X}) = U_{Total}(T, Q, Age).$$

2. For any PFS time, larger $Q =$ TTB decreases the utility more for an older patient than for a younger patient.

3. The utility $U_{PFS}(T)$ does not vary with age.

4. When there is no toxicity (TTB=0),
$U_{Total}(T, 0, Age) = U_{PFS}(T)$ since $U_{TTB}(0, Age) = 1$.

Elicited Numerical Utilities $U(PFS, TTB, Age)$

Age = 40 years

TTB (q)	PFS (t)				
	0	12	24	48	∞
0.000	0	25	50	95	100
0.025	0	25	50	95	100
0.050	0	23	49	95	100
0.100	0	20	48	93	100
0.150	0	17	44	90	100
0.500	0	5	32	80	100
1.000	0	0	0	0	0

Age = 50 years

PFS (t)				
0	12	24	48	∞
0	25	50	95	100
0	23	49	95	100
0	20	48	93	100
0	17	44	90	100
0	5	32	80	100
0	0	0	0	0
0	0	0	0	0

Age = 65 years

TTB (q)	PFS (t)				
	0	12	24	48	∞
0.000	0	25	50	95	100
0.025	0	20	48	93	100
0.050	0	17	44	90	100
0.100	0	5	32	80	100
0.150	0	0	0	0	0
0.500	0	0	0	0	0
1.000	0	0	0	0	0

Age = 85 years

PFS (t)				
0	12	24	48	∞
0	25	50	95	100
0	17	44	90	100
0	5	32	80	100
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Example: A **60-year-old patient** versus a **40-year-old patient**

Suppose $TTB = Q = 0.50$ and $U_{PFS}(36) = 80$.

For a 60-year-old patient: $U_{TTB}(0.50, Age = 60) = 0.50 \Longrightarrow$

$$U_{Total}(PFS, .50, Age) = U_{Total}(36, .50, 60) = U_{PFS}(36) \times U_{TTB}(.50, 60)$$

$$= 80 \times .50 = 40$$

For a 40-year old patient: $U_{TTB}(0.50, Age = 40) = 0.70 \Longrightarrow$

$$U_{Total}(PFS, .50, Age) = U_{Total}(36, 0.50, 40)$$

$$= 80 \times .70 = 56$$

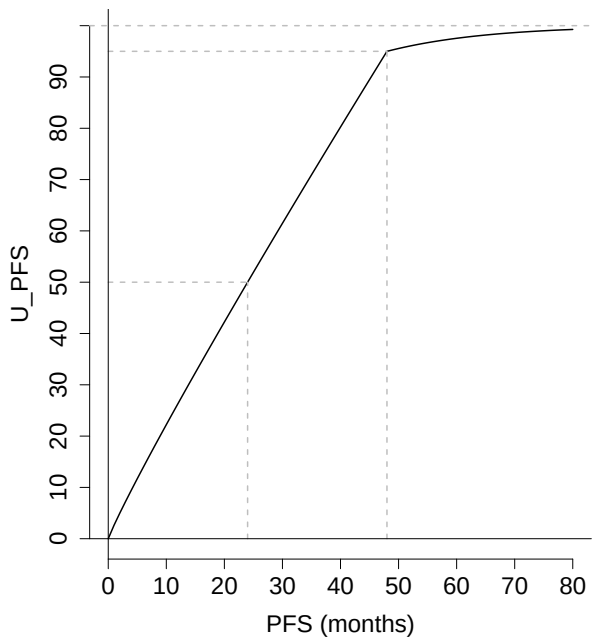
.

Specifying a Functional Form for U_{PFS}

Msaouel and Lim specified $U_{PFS}(24) = 50$, $U_{PFS}(48) = 95$, and $\lim_{t \to \infty} U_{PFS}(t) = 100$, and constructed the function

$$U_{PFS}(t) = \begin{cases} 95 \left(\frac{t}{48}\right)^a & \text{if } t < 48 \\ \\ \frac{100}{1+\exp(-b_1 t)} & \text{if } t \geq 48. \end{cases}$$

Fitting this to the 2 elicited values gave $a = 0.926$ and $b_1 = 0.061$.



Specifying a Functional Form for U_{TTB}

The multiplicative penalty term $U_{TTB}(Q, Age)$ was defined to decrease faster for older Age.

Msaouel and Lim established numerical values of $0 < U_{TTB} < 1$ for (Q, Age) pairs on a grid, and approximated them by the function

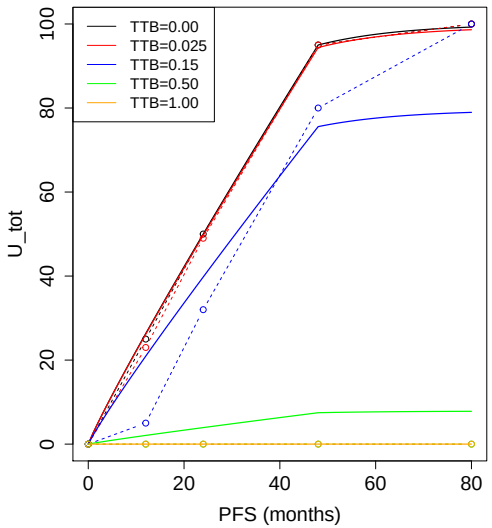
$$U_{TTB}(Q, Age) = \exp\{-Q^2/(2g^2(Age))\}, \text{ for } 0 \leq q \leq 1,$$

where

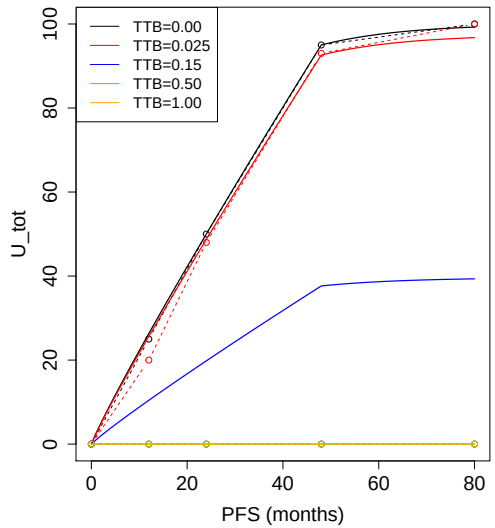
$$g(Age) = \exp(.823 - .05Age)$$

from the elicited numerical utilities

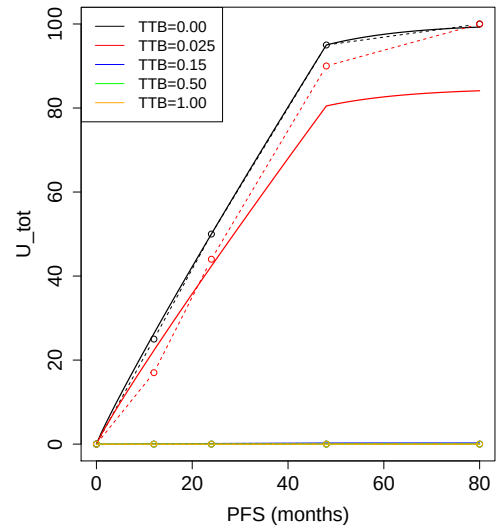
Elements of the PUFF for Age = 50, 65, and 85



U_{Total} for $Age = 50$



U_{Total} for $Age = 65$



U_{Total} for $Age = 85$

Using the Utility for Personalized Treatment Selection

Recall $\tilde{T} = \log(T)$. A new patient with prognostic covariates $\boldsymbol{X}^{new}$ treated with τ has **predictive outcome distribution**

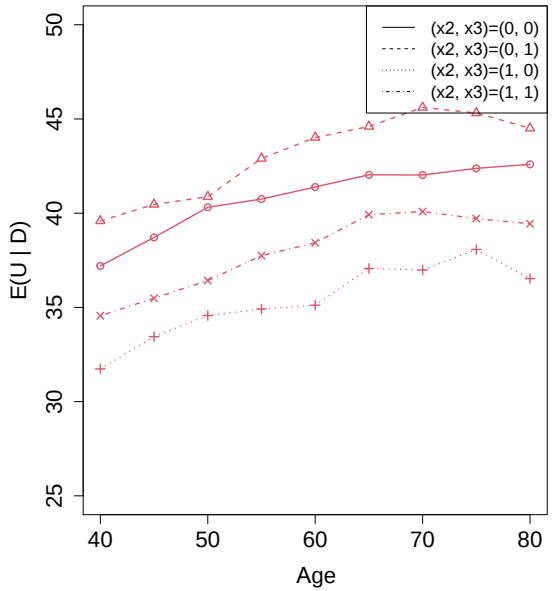
$$p(\tilde{T}, \boldsymbol{Z} \mid \tau, \boldsymbol{X}^{new}, \mathcal{D}).$$

The Predictive Mean Total Utility of treatment τ for a new patient with covariates $\boldsymbol{X}^{new}$ is

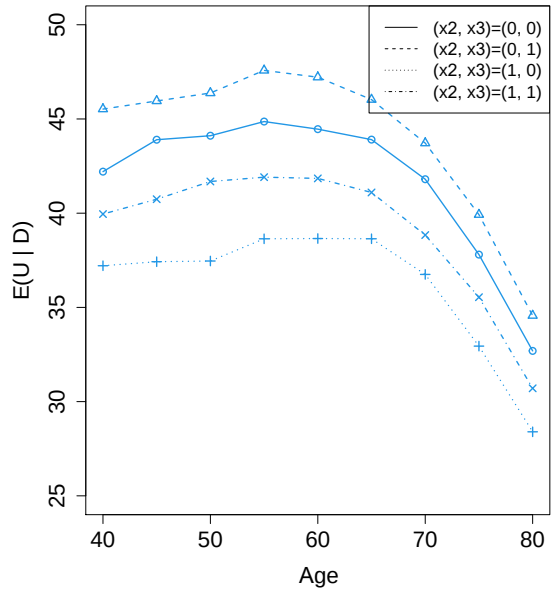
$$u_{Total}(\tau, \boldsymbol{X}^{new}, \mathcal{D}) =$$

$$\sum_{z_1=0}^{5} \dots \sum_{z_{21}=0}^{5} \int_{\mathbb{R}} U_{Total}(t, Q(z), Age^{new}) p(\tilde{t}, z \mid \tau, \boldsymbol{X}^{new}, \mathcal{D}) d\tilde{t}.$$

Decision for the new patient: Choose the treatment $\tau = L + B$ or L having larger $u_{Total}(\tau, \boldsymbol{X}^{new}, \mathcal{D})$.



$u_{Total}(L, \boldsymbol{X}^{new}, \mathcal{D})$



$\bar{u}_{Total}(L + B, \boldsymbol{X}^{new}, \mathcal{D})$

Posterior predictive mean utilities for different values of $\boldsymbol{X}^{new}$ = (Age, I[Active Disease], I[$>$ 24 months w/o prior disease]) for L = *Letrozole* and $L + B$ = Letrozole + Bevacizumab

A Predictive Criterion for Choosing a Treatment Based on X

Posterior predictive probability that $L + B$ has larger total utility than L for a new patient with prognostic covariates $\boldsymbol{X}^{new}$:

$$\Delta(\boldsymbol{X}^{new}) = \Delta(Age^{new}) =$$

$$\Pr\left\{ U_{Total}(T, Q, Age^{new}, L + B) > U_{Total}(T, Q, Age^{new}, L) \,\middle|\, \mathcal{D} \right\}$$

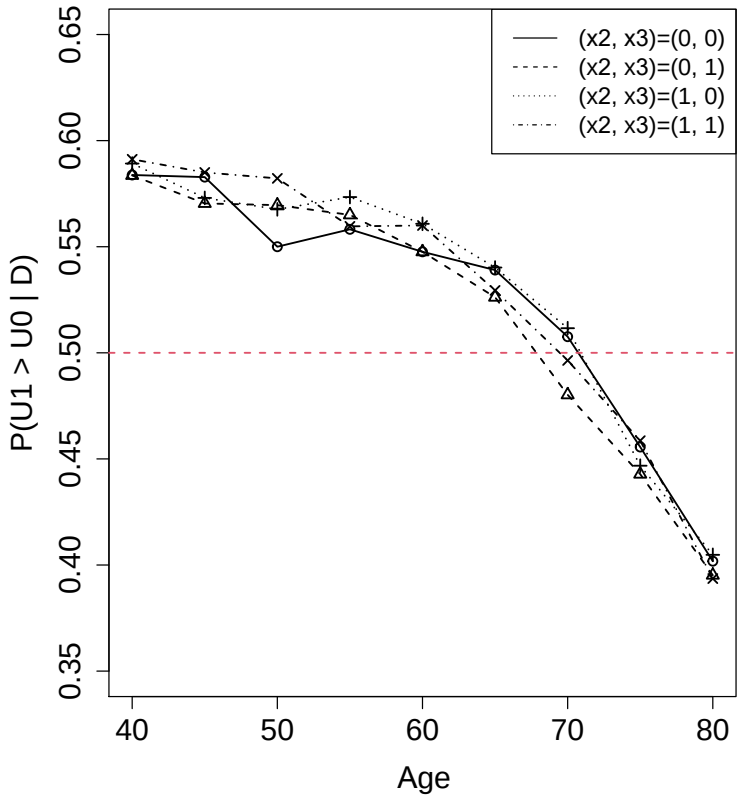
Interpretation

$\Delta(\boldsymbol{X}^{new}) > .50 \Longrightarrow L + B$ is more desirable than L.

$\Delta(\boldsymbol{X}^{new}) = .50 \Longrightarrow L + B$ and L are equally desirable.

$\Delta(\boldsymbol{X}^{new}) < .50 \Longrightarrow L$ is more desirable than $L + B$.

A Solution to the Decision Problem



$\Delta(X^{new})$

$L + B$ **is preferable if** $Age < 70$ **and** L **is preferable if** $Age \geq 70$

A computer program **utility-analysis** is available from
https://users.soe.ucsc.edu/juheelee/

References

Dickler, et al. Phase III trial evaluating letrozole as first-line endocrine therapy with or without bevacizumab for the treatment of postmenopausal women with hormone receptor positive advanced-stage breast cancer: CALGB 40503 (alliance). *J Clinical Oncology*, 34(22): 2602, 2016

Lee J, Thall PF, Lim B, Msaouel P. Utility based Bayesian personalized treatment selection for advanced breast cancer. *J Royal Statistical Society, Series C*. 71:1605-1622, 2022.

Muller P, Quintana A, Jara A, and Hanson T. *Bayesian Nonparametric Data Analysis*. Berlin, Springer-Verlag, 2015.

Thall, PF. *Bayesian Precision Medicine*. Chapman & Hall/CRC Press, 2024.

Required Reading