Product Lifecycle Services (PLS), NBS

A Semi-parametric Bayesian Model for Personalized Medicine

Ayan Das Gupta, Lead Statistician,
PLS, Novartis Business Services

PhUSE SDE 2018, Delhi
March, 2018

NOVARTIS

Disclaimer

- This presentation is based on publicly available information (including data relating to non-Novartis products or approaches).
- The views presented are the views of the presenter, not necessarily those of Novartis.
- These slides are intended for educational purposes only and for the personal use of the audience. These slides are not intended for wider distribution outside the intended purpose without presenter approval
- The content of this slide deck is accurate to the best of the presenter’s knowledge at the time of production.

Index

Introduction

NOVARTIS

Personalized Medicine

Preventative and predictive way to manage a better healthcare

- Utilizes patient's virtual cloud of data points – genetics, medical history, demographics, environment, disease type, etc.
- Demystify disease and population subtype characteristics
- Potentially predict risk factors, and treatment outcomes
- Help clinicians to make right decisions about patients

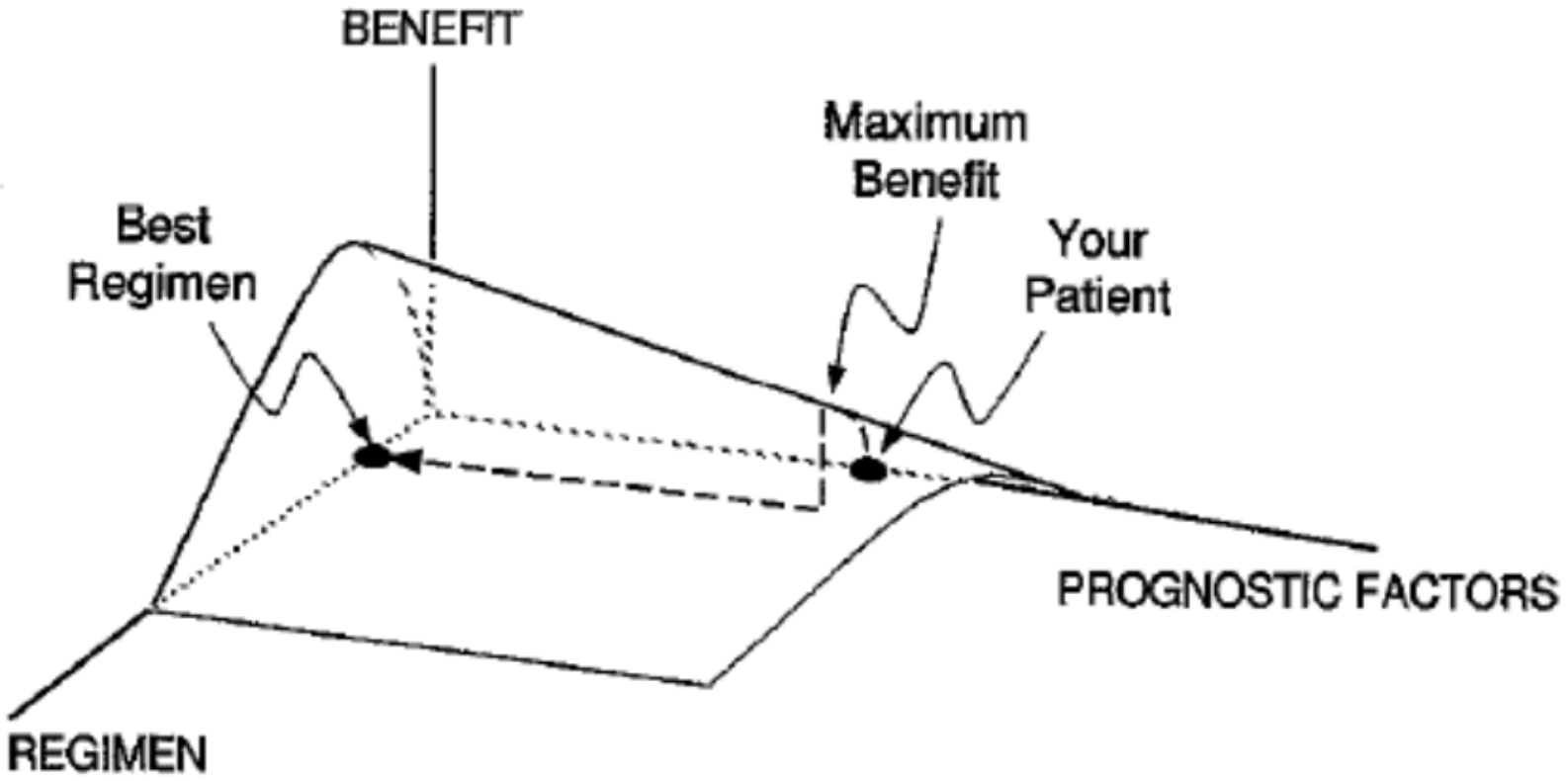
Adaptive Trial Designs for Personalized Therapy

- Adaptive trial designs which consider many therapies in the same trial, and using accruing data and **Bayesian techniques** to gradually matching patient subtypes with their most appropriate therapies.
- *"Improved utilization of adaptive and Bayesian methods" could help resolve low success rate of and expense of phase 3 clinical trials* - Janet Woodcock, Director CDER FDA
- FDA's Critical Path Opportunities Report (2006) – *"Uncovered a consensus that the two most important areas for improving medical product development are biomarker development and streamlining clinical trials."*

Innovations in Trial Designs - Use of Bayesian Techniques

- **I-SPY 2 –** In 2010, the Biomarkers Consortium initiated a groundbreaking adaptive biomarker-driven trial in neo-adjuvant breast cancer
  - Identify improved treatment regimens for subsets on the basis of molecular characteristics (biomarker signatures) of their disease
  - Uses Bayesian predictive probability to 'graduate' or 'drop-off' any therapy from the trial with their corresponding biomarker signature(s)
- **ADAPT- IT** – Funded jointly by the NIH and FDA to bring innovative trial design to the NETT (Neurological Emergencies Treatment Trials) network.
  - The ADAPT-IT teams explored and developed confirmatory adaptive clinical trial designs with the individual trial teams of NETT (five teams).

Therapeutic Response Surface – An Example (Clin Pharmacol Ther 1997)



Role of Bayesian Techniques in Personalized Medicine

- Efficiently *utilizes* the information to *provide* the disease sub populations with the ‘best possible’ regimen thereby *protecting* patients from getting over-treated with ineffective and excessively toxic regimen.
- To develop innovative clinical trial **designs**
  – Monitoring strategies, improved dose-finding strategies, adaptive randomization, and improved screening designs
- To develop innovative **methods** to analyze data from collaborative research studies
  – Data mining, adaptive learning, and integrating information across multiple platforms

Objective

NOVARTIS

Objectives

➢ Our Objective

- To build a semi-parametric Bayesian model for personalized medicine

➢Ultimate Clinical Objective

- Guide clinicians to recommend the most effective treatment to patients of certain age and belonging to distinctly different disease subtypes
- Monitor patient's response over the years and utilizing it for dynamic decisions

Data Description

NOVARTIS

Description of the data

➢The data comprises of:

- Response
  – **Treatment Response** – '*y*' (continuous, lower the better)
- Predictors
  – **Age** – '*x*' (in years)
  – **Treatment** – 0 => Standard Treatment, 1 => New Treatment
  – **Population/ Disease Subtype** – 0, 1 and 2

➢We used *simulated* data for our problem, although real data from various historical clinical trials, comprises of similar predictors, may also be analyzed.

Data Simulation

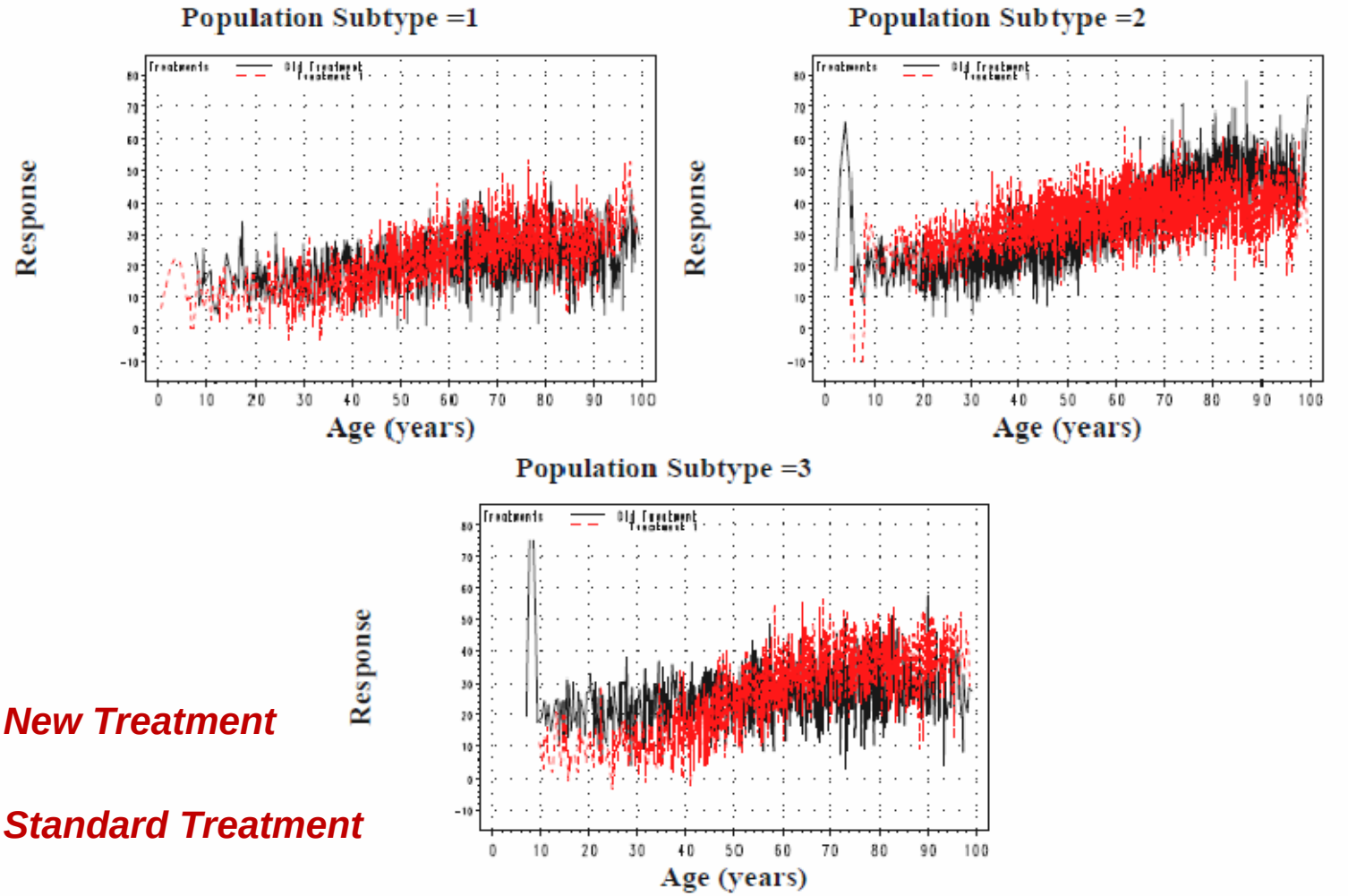
- The data were split into a ***training*** set (n_1 =10,000) and a ***test*** set (n_2 =1,000).
- The response and predictor variables for each sample dataset are generated as follows
  - $Treatment_k \sim Bernoulli\ (p_{trt})$ *and* $Type_k \sim Binomial\ (2, p_{type})$, where the subscript *k* denotes each sample unit in the dataset.
- Both the *success probabilities* are considered to be 0.5, which necessitates approx. 1:1 sample proportion for treatments and 1:2:1 sample proportion for population subtypes
  - $Age_k \sim 100 * Beta\ (3, 2)$
  - $Response_k \sim Normal\ (\mu, 5 + Age/25)$
  - $Mean\ respone\ \mu \sim f(Age, Type, Treatment)$

Data Simulation (continued)

- For each choices of *(Type=i, Treatment=j)*, the mean response ***μ*** has the logistic form $\mu = a_{ij} + b_{ij} * logit\{(Age - c_{ij})/d_{ij}\}$, where $logit\ (x) = e^x/(1+e^x)$
- The choices of constants *a, b, c,* and *d*, are chosen such that there exist **different treatment behaviours** over **different population subtypes**.

(Type, Treatment)	(a, b)	(c, d)
(0, 0)	(15, 10)	(50, 10)
(0, 1)	(10, 20)	(50, 10)
(1, 0)	(20, 30)	(60, 10)
(1, 1)	(25, 15)	(60, 10)
(2, 0)	(20, 10)	(50, 10)
(2, 1)	(8, 30)	(50, 10)

Figure I: Comparison of True Response over Subtypes – Training data



Methodology

NOVARTIS

Statistical model (parametric)

- A general form of the semi-parametric model is considered

$$Y = f_{param}(age, type, treatment; \beta_k) + f_{semiparam}(age, \gamma_j) + \varepsilon$$

where ε is the unexplained (random error) component of the model

- Two linear choices of the ***parametric*** part of the functional form, depending on ***with*** or ***without*** interaction terms with ***age***, are considered as follows:

$$f_{param}(age, type, treatment; \beta_k) = \beta_0 + \beta_1 * treatment + \beta_2 * type + \beta_3 * treatment * type + \beta_4 * age * type + \beta_5 * treatment * age + \beta_6 * treatment * type * age \ldots\ldots\ldots (1)$$

$$f_{param}(age, type, treatment; \beta_k) = \beta_0 + \beta_1 * treatment + \beta_2 * type + \beta_3 * treatment * type \ldots\ldots\ldots (2)$$

Statistical model (semi-parametric)

- The unknown, ***semi-parametric*** functional form can be approximated either by a *cubic B-spline* (with varying number of knots) or by a higher order *polynomial*

- Cubic B-Spline Approximation: The following model is selected, with varied number of knots 'j' ϵ {5, 10, 15, 20, 25, etc}

$$f_{semiparam}(age;\ \gamma_j) = \sum_{j=1}^{J} B_j(age) * \gamma_j \ldots\ldots (3)$$

- Polynomial Approximation: The following (cubic and quadratic) models are selected

$$f_{semiparam}(age;\ \gamma_j) = \gamma_1 * age + \gamma_2 * age^2 + \gamma_3 * age^3 \ldots . (4)$$

$$f_{semiparam}(age;\ \gamma_j) = \gamma_1 * age + \gamma_2 * age^2 \ldots . (5)$$

Statistical model (continued)

- Based on the judgment on performances of predicting the test data set, the following models are shortlisted

Parametric part	Semi-parametric part	Number of Knots	Acronym Used
With Interactions of Age	Cubic B-Spline Model	5, 10, 15, 20, and 25	Spline-5, Spline-10, etc.
	Cubic polynomial Model	N.A.	Cubic
	Quadratic polynomial Model	N.A.	Quad
Without Interactions of Age	Cubic B-Spline Model	5, and 10	Spline-5A, Spline-10A
	Cubic polynomial Model	N.A.	Cubic-A
	Quadratic polynomial Model	N.A.	Quad-A

Prior Selection

- Flat priors for intercept term β_0 and Gaussian prior on random error component $\varepsilon \sim Normal\ (0, 1/\tau_e)$
- Independent Cauchy (t with ν=1) priors for β_k and γ_j, which is of the form $t_\nu(0, 1/\lambda)$, where λ is the scaling parameter
- In hierarchical terms, the above Cauchy priors can be reduced to

$$\beta_k \sim N\left(0, \frac{1}{\tau_{\beta_k}}\right) where\ \tau_{\beta_k} \sim G\left\{\frac{\nu}{2}, \left(\frac{\nu * \lambda}{2}\right)\right\}$$

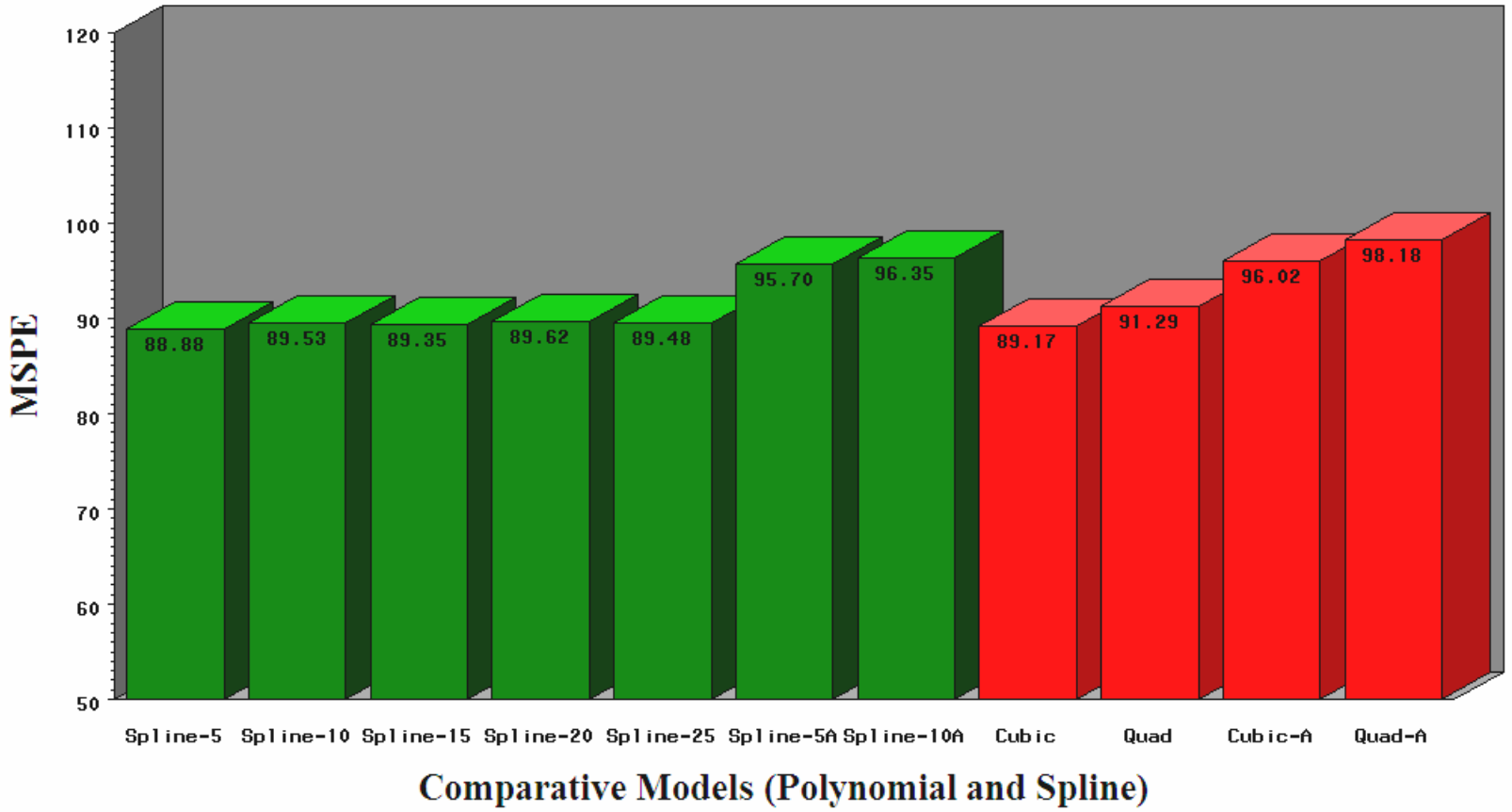
$$\gamma_j \sim N\left(0, \frac{1}{\tau_{\gamma_j}}\right) where\ \tau_{\gamma_j} \sim G\left\{\frac{\nu}{2}, \left(\frac{\nu * \lambda}{2}\right)\right\}$$

- The hyper priors for λ and τ_e are taken as G (a_1, b_1) and G (a_2, b_2) respectively
- Small values of a_1, b_1, a_2, and b_2 (= 0.01) are chosen, for non-informative choices

Model fit

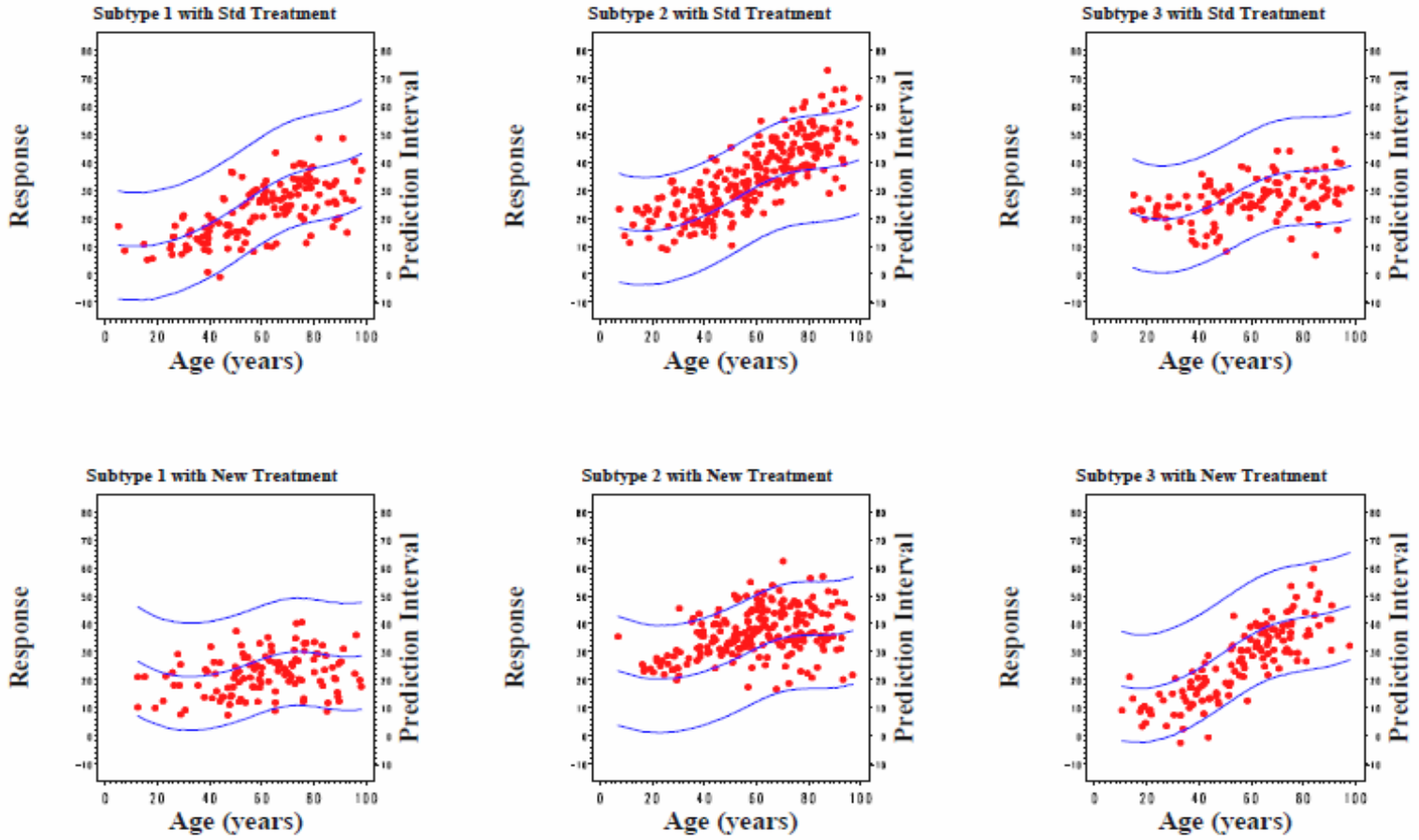
- The final shortlisted models are fitted to the training data and validated on the test set using posterior predicted response
- Models are compared using Mean Squared Prediction Error (MSPE) and coverage of prediction intervals (based on HPD)
- The model which provides the least MSPE and better coverage (95% prediction interval) is declared to be the ‘best’ under our set-up
- Based on the analyses, Figure II, and Figure III, the **cubic B-Spline model with 5 knots under ‘presence of interaction terms’** is selected
- A detailed prior sensitivity result proves the robustness and supports the finding

Figure II: Mean Squared Prediction Error (MSPE) for various models



NOVARTIS

Figure III: 95% Prediction interval over cohorts – B Spline (5 knots)

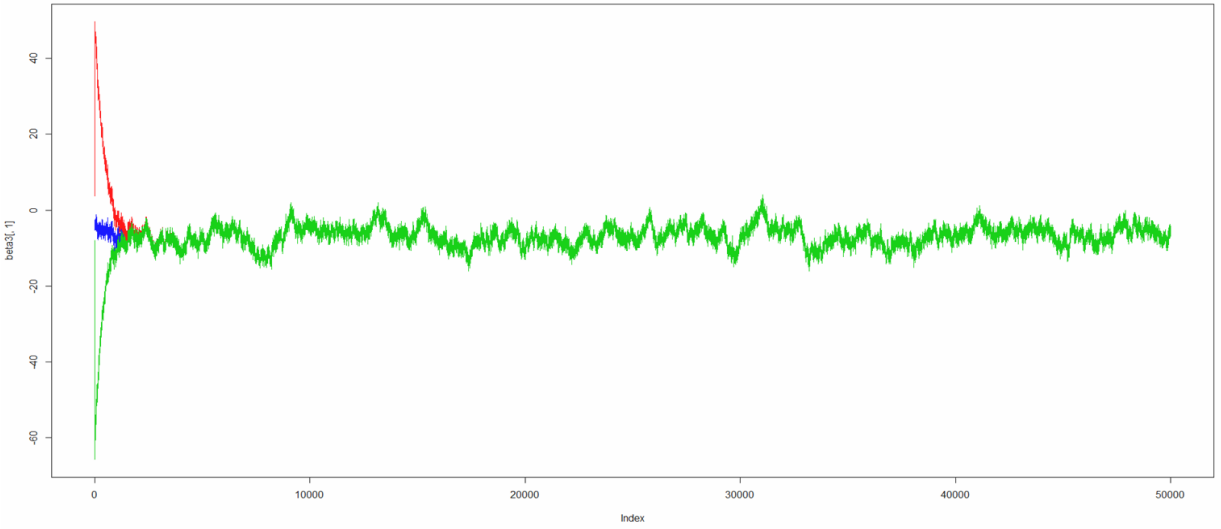


Computational Details

NOVARTIS

Computational details

- Random samples from posterior distributions were drawn using MCMC algorithm
- *Blocked Gibbs* sampler was applied, with β_0, coefficient vector, τ_e, and λ as four blocks
- Posterior statistics were estimated with **50,000** iterations and **1,000** burn-in samples
- The different chains are found to converge, proving the effectiveness of the algorithm
- The trace plot for β_3, using three independent chains with different initial values is shown below



Results

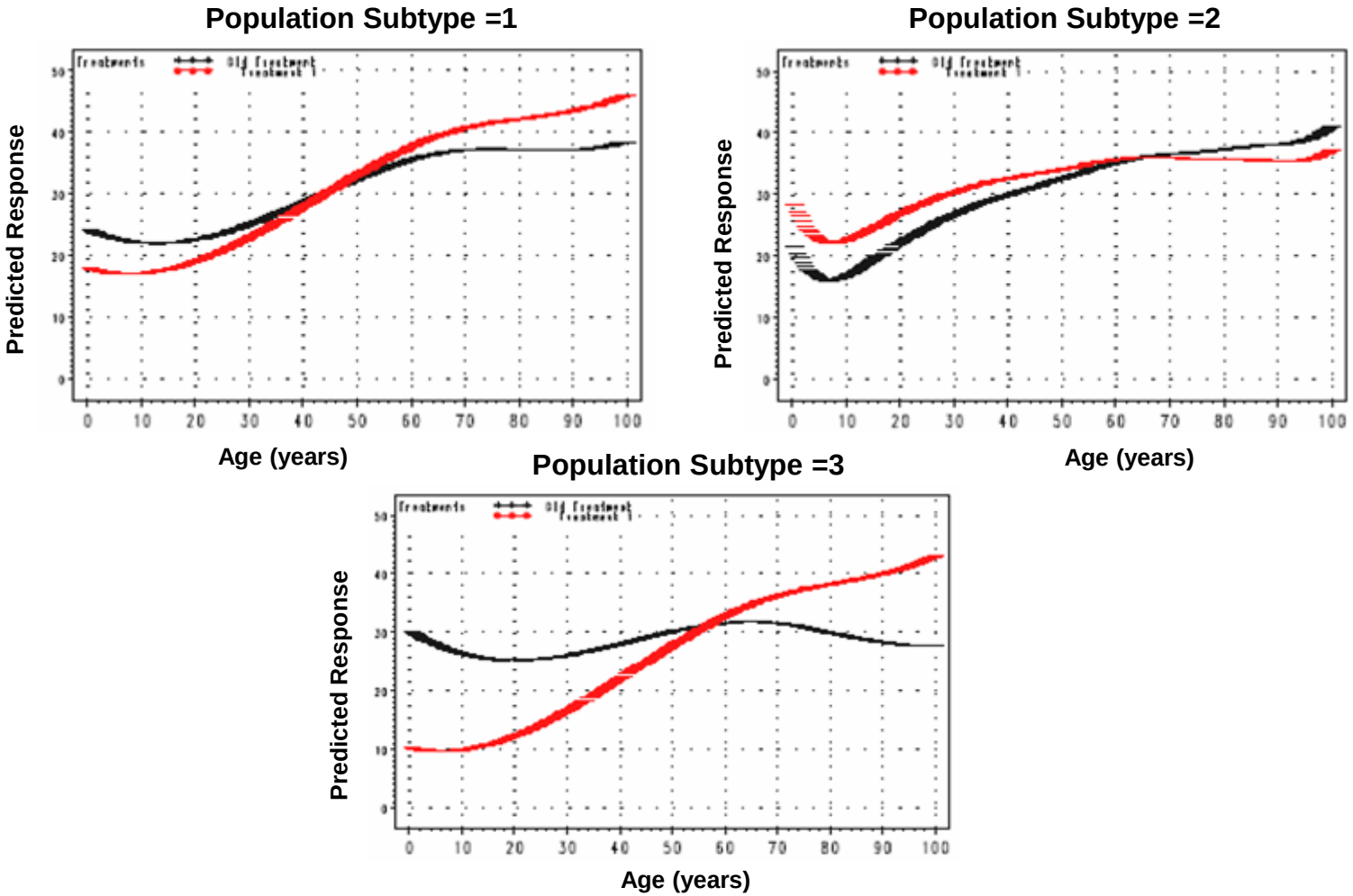
NOVARTIS

Posterior Summaries of Model Parameters

Parameters	Mean (s.d.)	Parameters	Mean (s.d.)
Intercept	10.58 (1.773)	Treatment*Age*Type	0.25 (0.013)
Treatment	19.97 (1.040)	Basis Spline parameter	-2.62 (2.588)
Type	7.00 (0.597)	Basis Spline parameter	6.59 (1.562)
Treatment*Type	-13.15 (0.852)	Basis Spline parameter	28.62 (2.022)
Age*Type	-0.09 (0.010)	Basis Spline parameter	27.82 (1.900)
Treatment*Age	-0.35 (0.016)	Basis Spline parameter	32.48 (2.239)
Error precision	0.01 (0.001)	Lamda λ	45.78 (40.861)

- Posterior estimate of the scale parameter λ is high – also sensitive to choices of priors and ν
- ***General Rule***: The effect of a model parameter can be considered to be insignificant if the posterior estimate is close to zero and with very high precision

Figure IV: Comparison of predicted response over population subtypes



Results aimed for medics

- The clinical behaviors in **Subtype I** are moderately different
  - New treatment promises to target younger patients (< 40 years), very less effective otherwise
  - Standard treatment is more effective for older patients, consistent performance for this group
- On the contrary, the clinical behaviors in **Subtype II** is almost similar!
  - New treatment fails to be more efficacious, except for very older patients (> 70 years)
  - Existing standard treatment over-performs the new one, though marginally
- The clinical behaviors of the two treatments in **Subtype III** is entirely different!
  - New treatment works exceedingly well, but only for patients less than 50 years of age
  - Performance of the standard treatment is consistent over age, recommended for older patients
- Overall, the new treatment fails to show consistent performance in older patients
- *The performance plots would highly benefit clinicians – helps to follow-up patients*

Prediction Intervals of Response for Treatment and Treatment Difference

➢ We consider *nine* individuals (young, middle, old) belonging to three different subtypes (I, II, and III)

Population Subtype	*Age*	*Standard Treatment*	*New Treatment*	*Difference (New – Std)*
Subtype I	20	(4.63, 36.38)	(1.08, 32.80)	(-4.74, -2.14)
Subtype I	50	(16.78, 48.30)	(17.43, 49.02)	(0.07, 1.39)
Subtype I	80	(26.37, 57.83)	(31.28, 62.77)	(4.14, 5.77)
Subtype II	20	(-0.33, 31.38)	(4.36, 36.11)	(4.03, 5.51)
Subtype II	50	(14.59, 45.80)	(16.10, 47.32)	(1.19, 1.93)
Subtype II	80	(26.91, 58.54)	(25.24, 56.84)	(-2.09, -1.14)
Subtype III	20	(7.26, 39.24)	(-5.70, 26.24)	(-14.15, -11.58)
Subtype III	50	(14.52, 45.96)	(12.09, 43.57)	(-2.98, -1.69)
Subtype III	80	(19.19, 50.67)	(27.37, 58.91)	(7.43, 9.07)

Prior Sensitivity

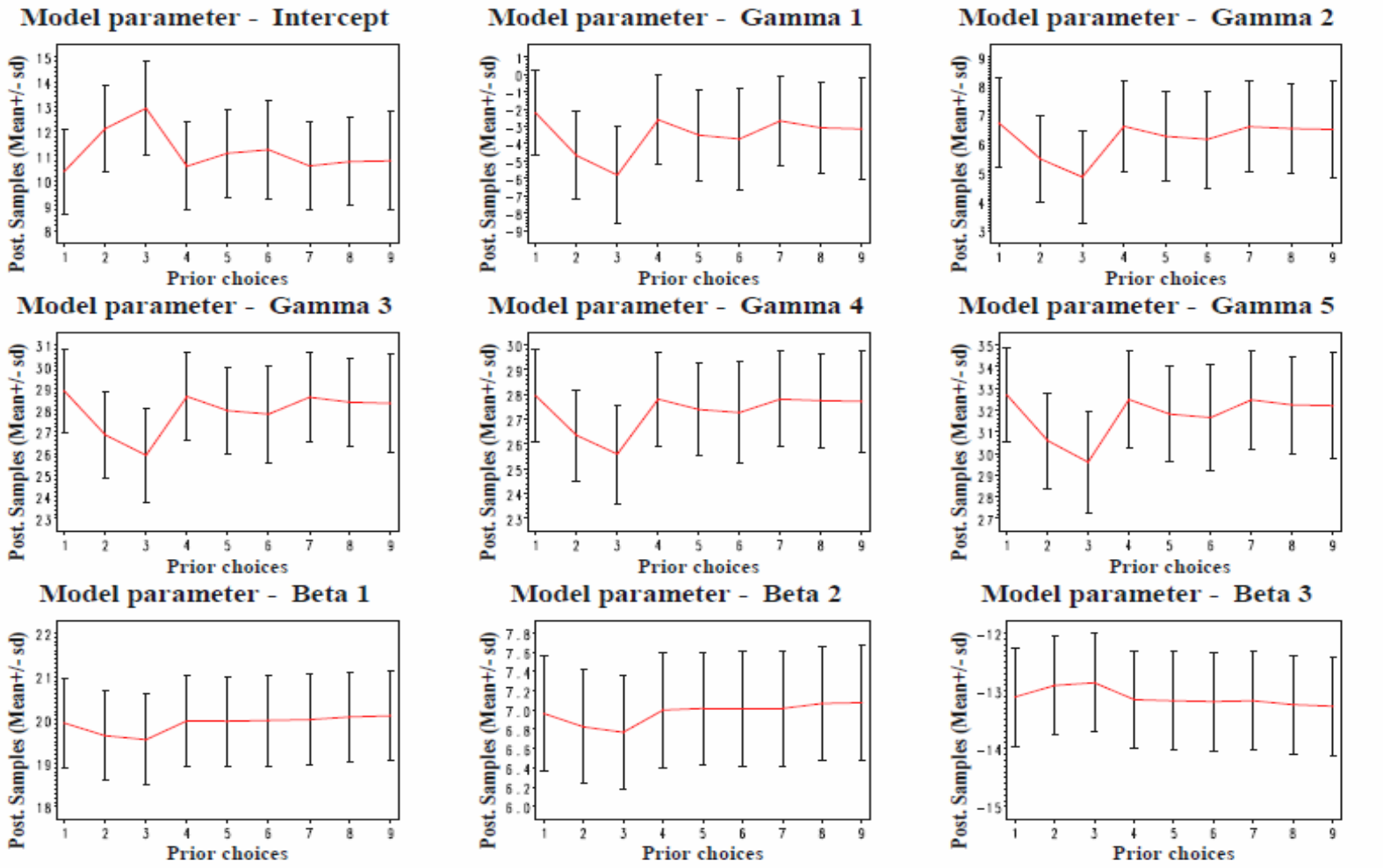
NOVARTIS

Prior Sensitivity

- To check any variations in model performance and posterior estimators
- *Three* choices of ν are considered, choices indicate the distributions – *Cauchy, t,* and *Normal*
- For each such choices of ν, *three* choices of $\mathbf{a_1}$, $\mathbf{b_1}$, $\mathbf{a_2}$, and $\mathbf{b_2}$ are chosen
- The choices on hyper priors indicate weak informative to non informative behaviours

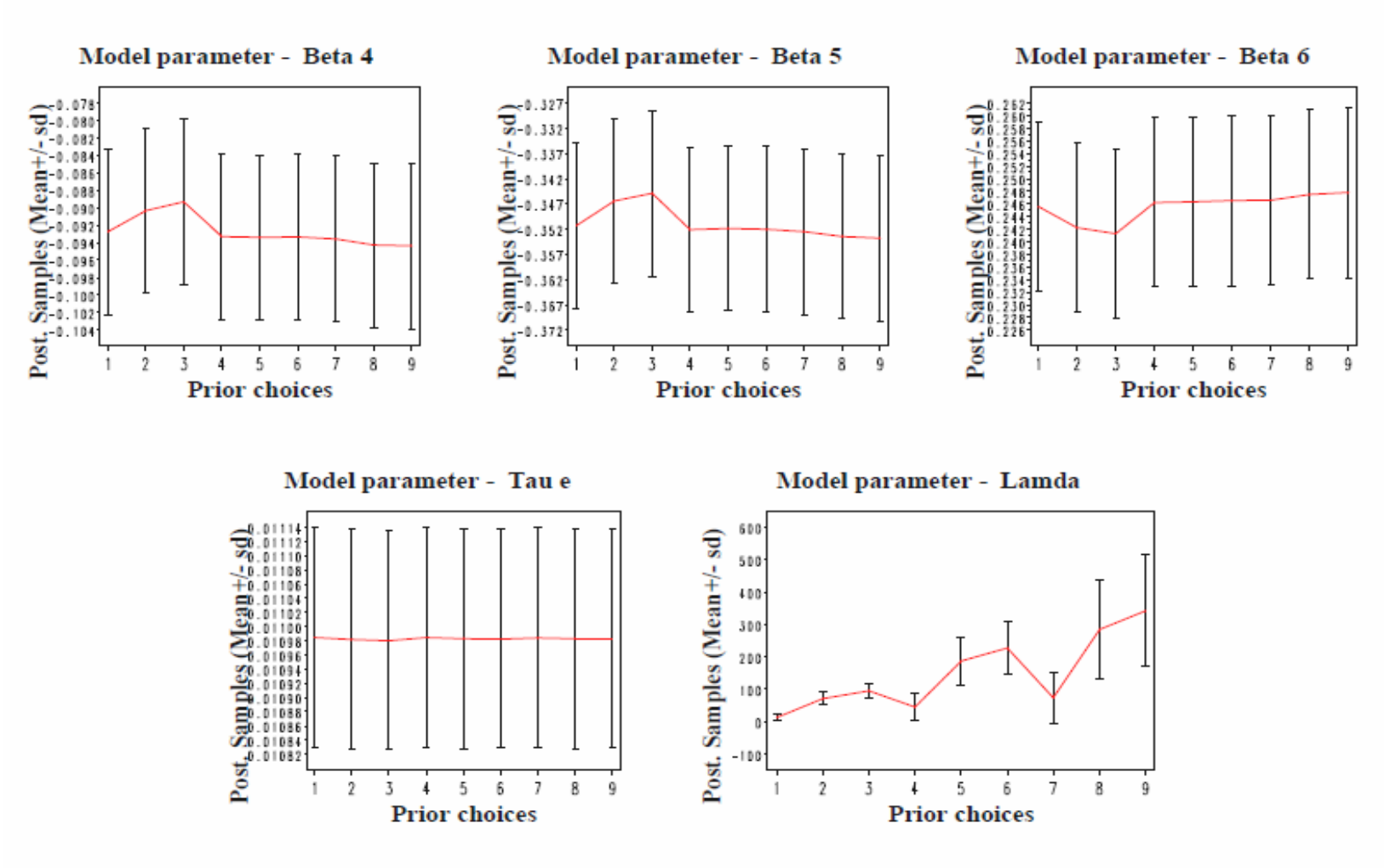
Choices of (a_1, b_1) and (a_2, b_2)	***Degrees of freedom (v)***	***Prior Index***	***Remarks***
(0.1, 0.1) (0.1, 0.1)	1, 10, and 100	1, 2, and 3	Weakly Informative
(0.01, 0.01) (0.01, 0.01)	1, 10, and 100	*4*, 5, and 6	Non Informative
(0.001, 0.001) (0.001, 0.001)	1, 10, and 100	7, 8, and 9	Most Non Informative

Figure V: Prior Sensitivity plot of Model parameters



NOVARTIS

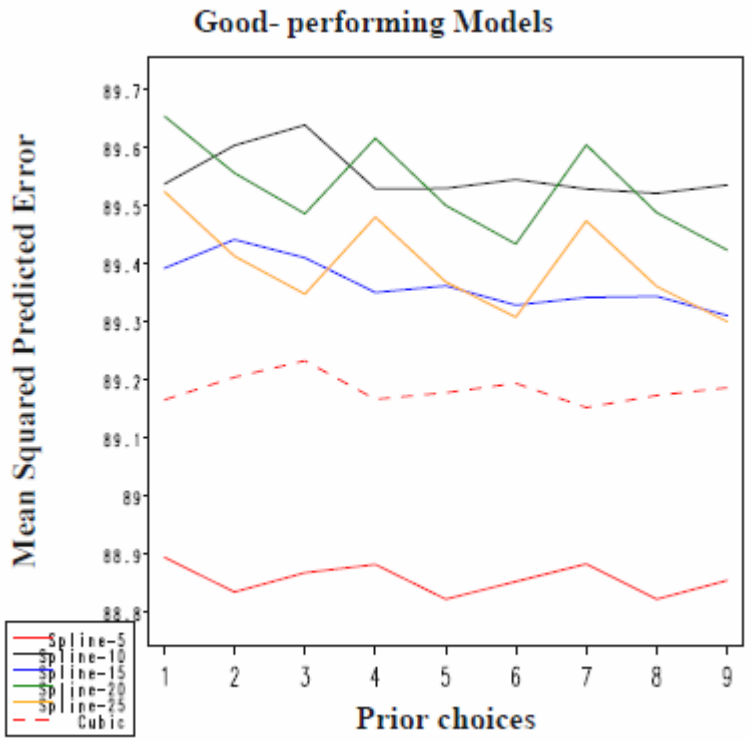
Figure VI: Prior Sensitivity plot of Model parameters

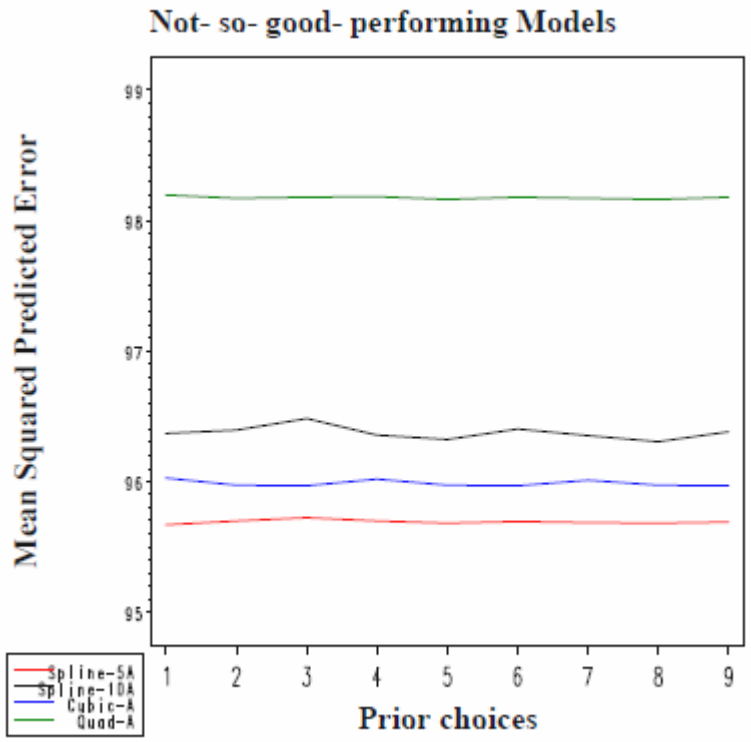


Prior Sensitivity

- The error precision remains unaffected
- The spline parameters show the same pattern and stabilizes with ‘Prior – 4’ onwards
- The ‘non-spline’ model parameters also stabilizes with ‘Prior – 4’ onwards
- The estimate (and variability) of scale parameter λ
  - Increases with increase in $\boldsymbol{\nu}$, with non-informative index ($\mathbf{a_1}$, $\mathbf{b_1}$) kept constant
  - Increases rapidly with increase in non-informative index, with $\boldsymbol{\nu}$ kept constant
- Justified to consider ‘Prior – 4’ in our model analyses

Figure VII: Prior Sensitivity plot of Model Diagnostics





- *The MSPE of the model 'Quad' is around **91.2**, which is between 'good' and 'not-so-good' models*
- The performance of the 'Spline-5' model is ***strongly robust*** under prior variations

NOVARTIS

Conclusion

NOVARTIS

Conclusion

- The Bayesian semi-parametric model, uses prior information about response/ factors, demystifies patient characteristics data, and helps clinicians to
  - Prevent patients by predicting risks and responses
  - Take right decisions to match individuals/ subtypes with best treatment regimens
  - Helps to monitor patients by dynamically varying treatments
- Protect patients getting over-treated with non-efficacious and more toxic drugs
- Helps to provide and maintain a better healthcare

Extensions

- In scenarios where more factors (genetics, safety, demographics, etc) are to be utilized, non-parametric Bayesian approaches like Dirichlet Process Mixture (DPM) models can be applied and compared with the semi-parametric cubic B-Spline models

Acknowledgments

- *Dr. Brian J. Reich*
- *Dr. Marie Davidian*

Thank you

NOVARTIS