A DeepNLME-based Tumor Growth Dynamics and Overall Survival Model for Non Small Cell Lung Cancer

Mohamed Tarek ♠ Lorenzo Contento ♠ Amit Roy ♠ Kiran Gauthier ♢ Anna Kondic ♢ Jun Shen ♢

Pumas-AI Inc. Dover, DE, USA ♠ Bristol-Myers Squibb, Princeton, NJ, USA ♢

1 Objective

To develop a deep nonlinear mixed effects (DeepNLME) tumor growth dynamics and overall survival (TGD-OS) model to identify and characterize patterns of non-small cell lung cancer (NSCLC) tumor response and the relationship to OS.

2 Methods

DeepPumas was used for model development.

2.1 Training and validation data split

- 6285 patients from 9 studies receiving immuno-oncology agents or chemotherapies.
- 8 studies were used for the training and 1 study was used for validation.

2.2 Tumor dynamics DeepPumas model

$$\text{SLD} \sim \text{Normal}(\mu, \sigma \cdot \sqrt{\text{NILSN}})$$
$$\mu(0) = \theta_1 \cdot e^{\eta_1}$$
$$\frac{d\mu}{dt} = \mu^{I(\bar{\mu}<0)} \cdot \bar{\mu}$$
$$\bar{\mu} = (\theta_2 + \eta_2) \cdot \text{NN}_1\left(\frac{\mu}{100}, \frac{t}{1000}, \eta_{ts}\right) + (\theta_3 + \eta_3)$$

- NILSN: the number of lesions
- $I(c)$ is an indicator function that is 1 when the condition c is true and 0 otherwise.
- $\mu' = \mu^{I(\bar{\mu}<0)} \cdot \bar{\mu}$ ensures the non-negativity of the ODE solution by construction. The reason is that the rate of change μ' will never be negative as $\mu \to 0$.

2.3 Jointly fitting the population parameters and random effects

- **Goal**: to fit the model's structure first to ensure that we can intentionally over-fit the data then reduce over-fitting using regularization.
- With enough regularization, we can avoid over-fitting. Hyper-parameter fine-tuned to maximize the log likelihood on the validation dataset.
- Training objective: maximize the joint probability of all the data and parameters, population parameters and random effects.

$$p(\theta) \cdot p(\eta \mid \theta) \cdot p(D \mid \theta, \eta)$$

- The random effects distribution will be mis-specified because we don't maximize the marginal likelihood (addressed in the next section).

2.4 Normalizing flow (NF) as the random effects' prior

Goal: learn an approximation of the true prior distribution of the random effects

$$p(\eta) = p(\eta; x) = \int p(\eta \mid y, \theta, x) \cdot q(x, y)\, dy$$

- q: true/empirical distribution of the subjects
- The NF is an invertible (neural network) function NN_2 of a standard Gaussian $z \sim \text{Normal}(0, I)$:

$$\eta = \text{NN}_2(z)$$

- Fitting steps:
  1. Sample from $q(x, y)$ by looping over each subject (x_i, y_i) from the population.
  2. Sample from the posterior distribution $p(\eta \mid y = y_i, \theta, x = x_i)$ using MCMC for each subject i.
  3. Combine the MCMC samples from all subjects.
  4. Fit the NF with maximum likelihood to the combined MCMC samples.
- The fitted NF is used as the new random effects' prior in the TGD model.

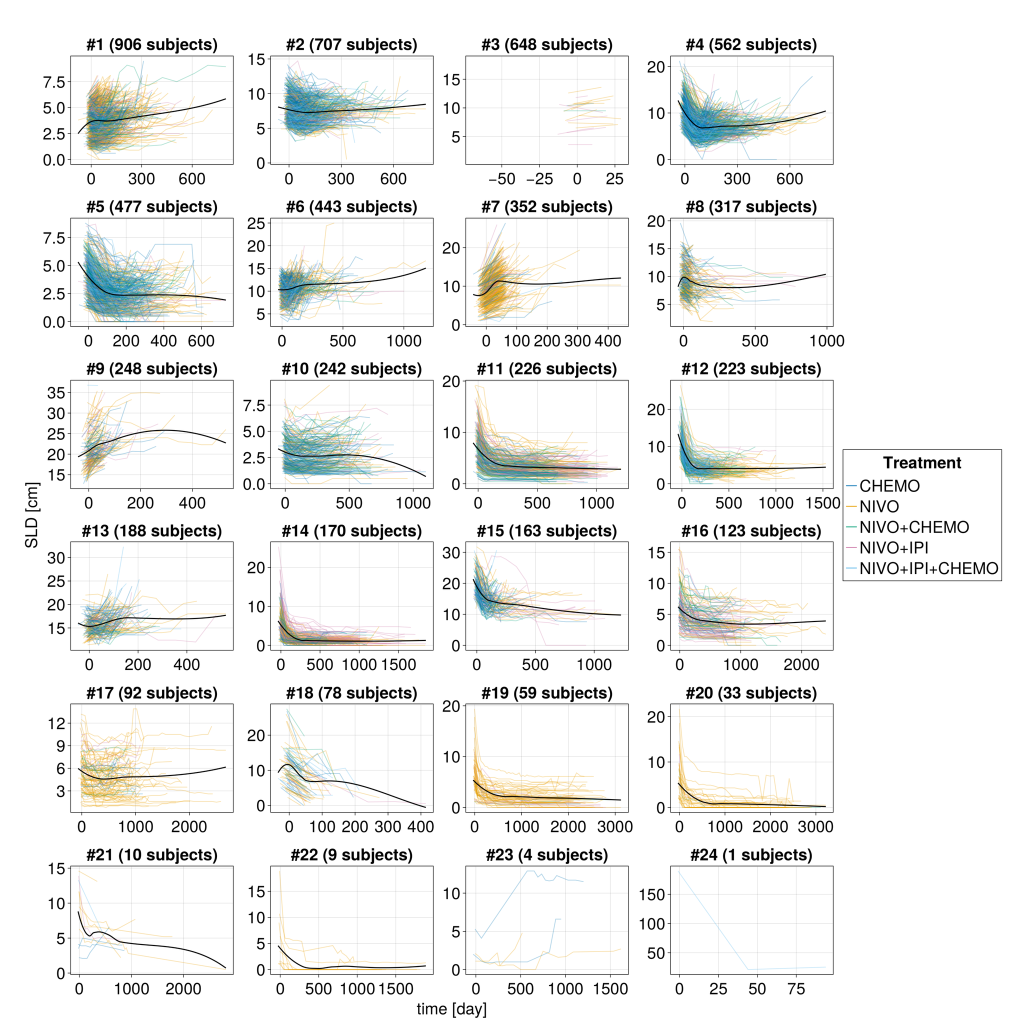
2.5 Overall survival DeepPumas model

$$\lambda = \lambda_0 \cdot \kappa \cdot (\lambda_0 \cdot t)^{\kappa-1} / (1 + \lambda_0 \cdot t)^{\kappa}$$
$$\lambda_0 = \lambda_1 \cdot exp(\beta' \cdot \text{xstd} + \text{NN}_3(\ldots))$$

- Log logistic hazard function with neural network NN_3 to extract features from the covariates and TGD model outputs to inform OS.
- xstd: standardized covariates.
- β vector of coefficients in the linear term.
- Inputs to NN_3: all baseline covariates (standardized), predicted SLD, PCSLD and rate of change of SLD from the TGD model, all scaled.

2.6 Model evaluation and data clustering

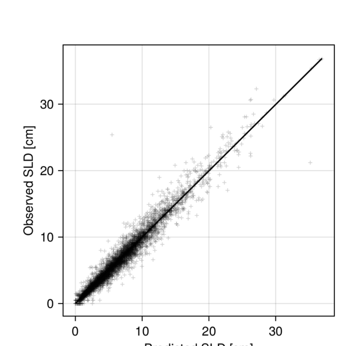
- Subjects were clustered into 24 clusters by their SLD profiles to understand the value of the DeepNLME model for different TGD profiles.

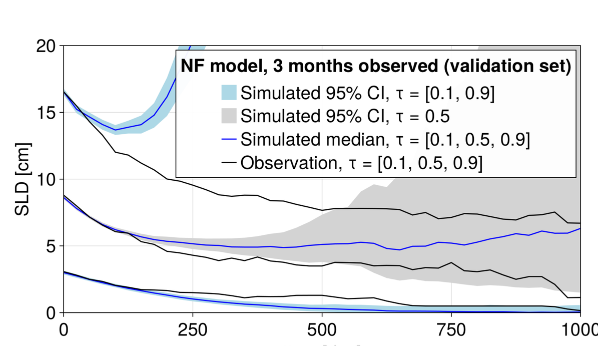


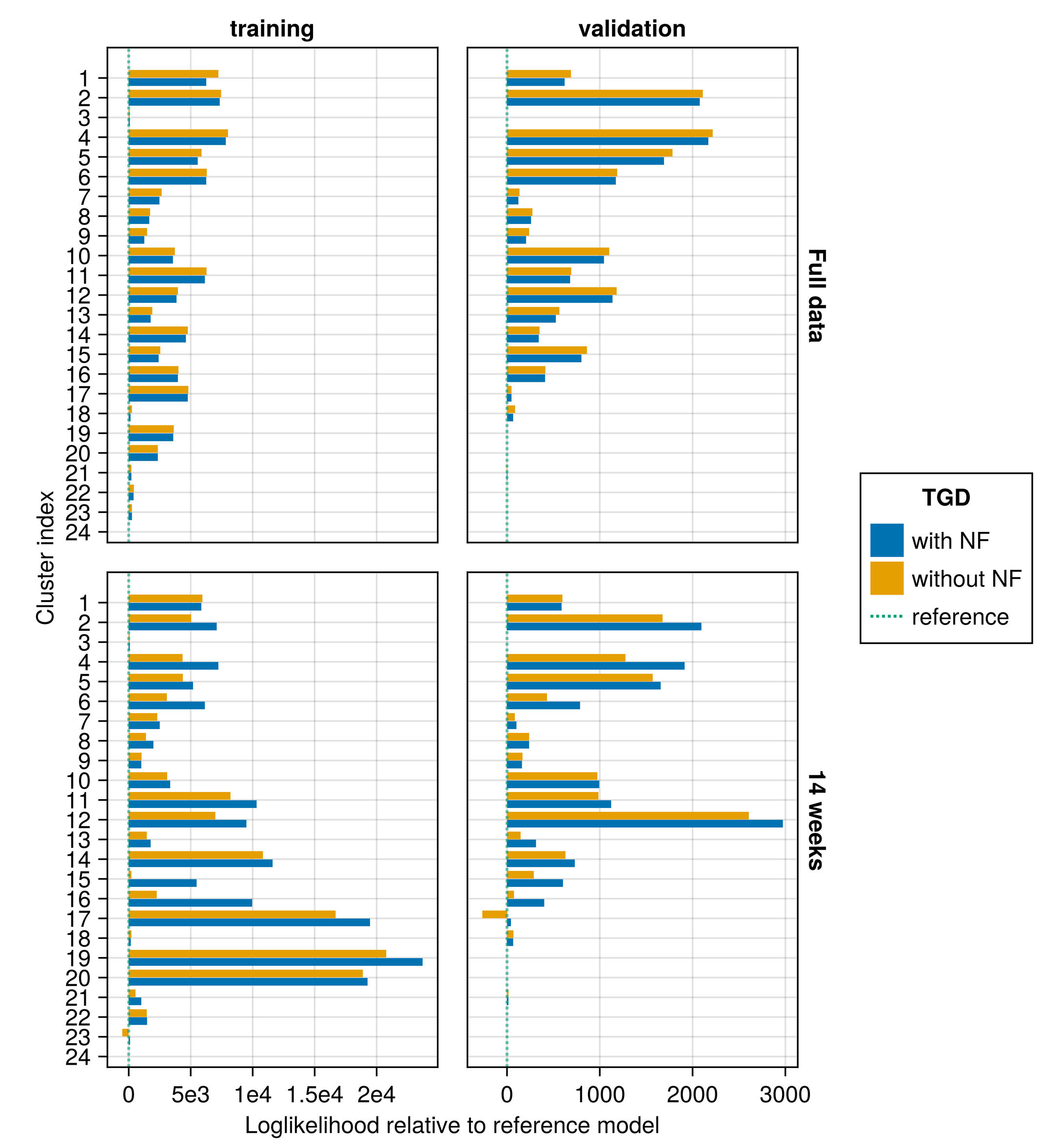
- The reference TGD-OS model consists of: 1) a mixture TGD model, and 2) a log logistic survival model with a linear covariates model.
- The total and per-cluster, training and validation, TGD and OS log likelihoods of the models were compared using the TGD empirical Bayes estimates (EBEs) given all the observations per subject versus only 14 weeks of data per subject.

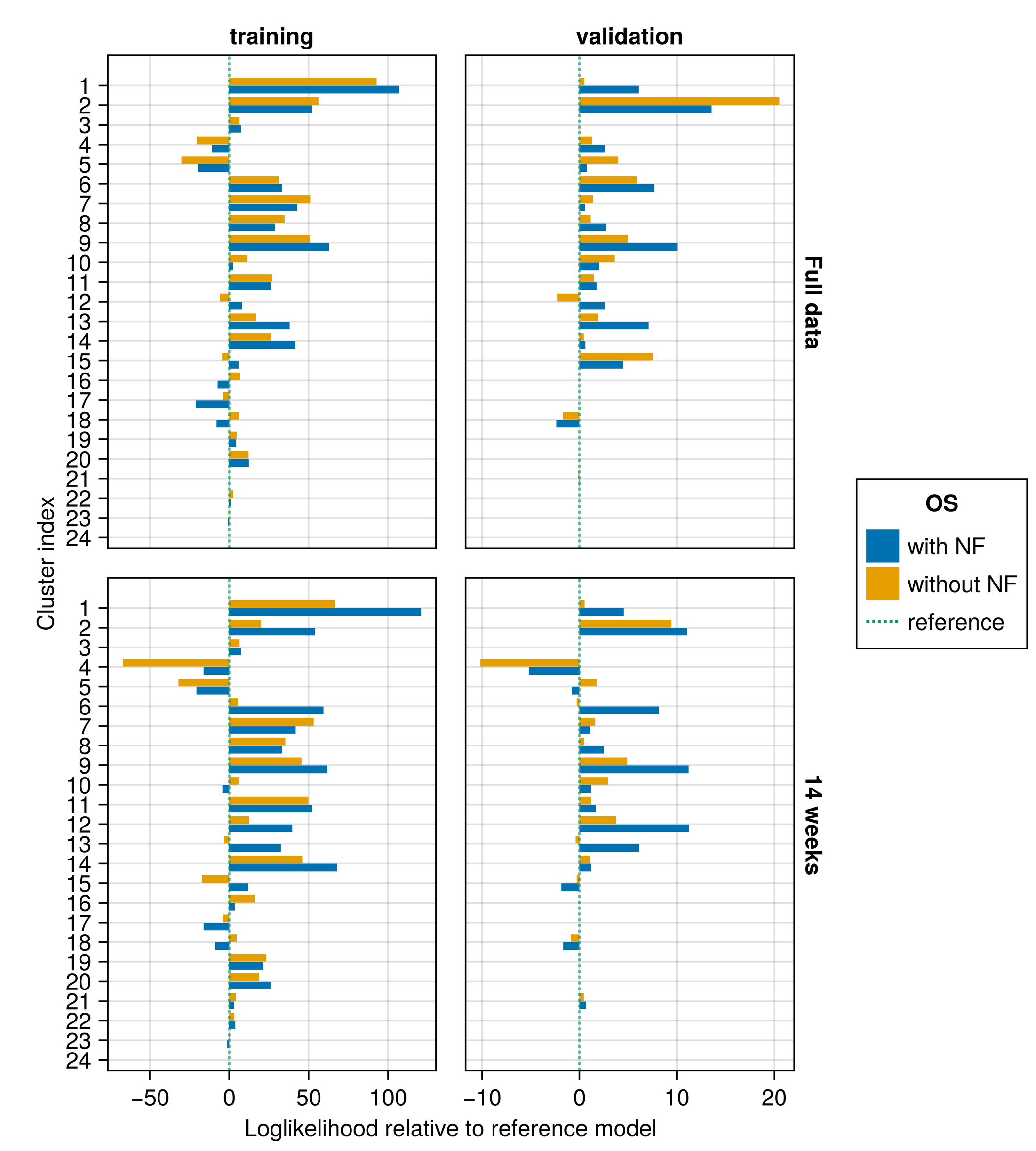
3 Results

The DeepNLME model had a higher OS log likelihood than the reference model by 868.3 points on the training data (8 studies) and 88.5 points on the validation data (1 study).









4 Discussion and Conclusion

- The long-term ipreds' accuracy was good with many observations per subject but limited with early data (e.g. 14 weeks).
- The DeepNLME model with NF has better TGD (OS) training *and* validation log likelihoods than the reference model across all (most) clusters using full data and 14 weeks' data to calculate the EBEs.
- The NF improves the training and validation, TGD and OS log likelihoods of the DeepNLME model with early data EBEs across almost all clusters.
- For clusters with conventional TGD profiles (e.g. 4 and 5), the reference model's OS log likelihood is better than the DeepNLME model's, likely due to the DeepNLME TGD model over-fitting the SLD measurements in these clusters misinforming the OS model.
- More work is needed to explore the value of DeepNLME TGD-OS models in decision-making given early data and to improve the models' prediction performance with such data.