A comparative benchmark study of empirical tumor size models in NSCLC clinical data

Anna Mishina*[1,2], Kirill Zhudenkov[(1-3)], Kirill Peskov[(1-3)]

M&S Decisions FZ-LLC, Dubai, UAE [1]; Research Center of Model-Informed Drug Development, I.M. Sechenov First Moscow State Medical University, Moscow, Russia [2]; Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences (RAS) [3]

Abstract

In this work, we conducted a systematic comparative analysis of frequently used tumor size dynamics models based on their descriptive and predictive ability validated against representative NSCLC clinical data.

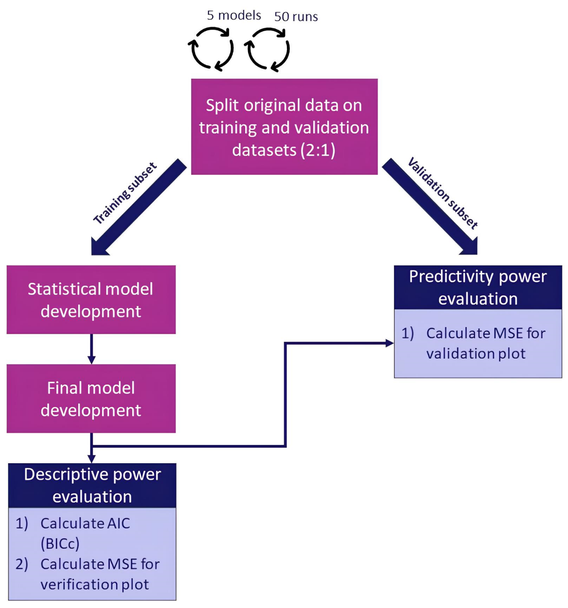
Introduction

In oncology, tumor size dynamics models provide the basis for delineating the progression of solid tumors over time. These models are further applied in developing joint [1-2] and sequential models [3] of longitudinal biomarkers and patient survival, improving insights into tumor growth and treatment outcomes. Additionally, the parameters of these models can be used as surrogate endpoints in clinical research [4]. **Despite the widespread application of these empirical models, the choice of the optimal one often lacks proper justification and still have unresolved methodological issues.**

Methods

We analyzed longitudinal sum of target lesions (SLD) data (RECIST 1.1 criteria [5]) from advanced non-small cell lung cancer (NSCLC) patients treated with EGFR tyrosine kinase inhibitor - Erlotinib (NCT00364351, 381 SLD profile), available at the ProjectDataSphere [6].

The scheme below represents a repeated cross-validation approach applied in current analysis. For each split between training and validation datasets (N=50), we performed unified model diagnostics (at training dataset) and validation (validation dataset).



In the current work the following 5 frequently used SLD empirical models were used [7].

TABLE 1. Empirical SLD dynamics models under investigation

Model name	Equation
Biexponential model (BiExp)	$SLD = Base \cdot (e^{k_g \cdot t} + e^{-k_d \cdot t} - 1)$
Linear exponential model (LExp)	$SLD = Base \cdot e^{-k_d \cdot t} + k_g \cdot t$
Quadratic exponential model (QExp)	$SLD = Base \cdot e^{-k_d \cdot t} + k_{g1} \cdot t + k_{g2} \cdot t^2$
Biexponential with sensitivity parameter model (BiExpS)	$SLD = Base \cdot (\varphi \cdot e^{-k_d \cdot t} + [e^{k_g \cdot t} - \varphi])$
Tumor growth inhibition model (TGI)	$SLD = Base \cdot e^{k_g \cdot t - \left(\frac{k_d}{\lambda}\right) \cdot (1 - e^{-\lambda \cdot t})}$

k_g, k_{g1}, k_{g2} - growth rate constants; k_d - shrinkage rate constant of tumor due to applied cancer treatment; φ - sensitive fraction of the tumor; λ - treatment efficacy decay rate constant.

Model development followed standard population model methodology [8]. Nonlinear mixed effects approach was performed in R using the lixoftConnectors (2023.1) API for Monolix.

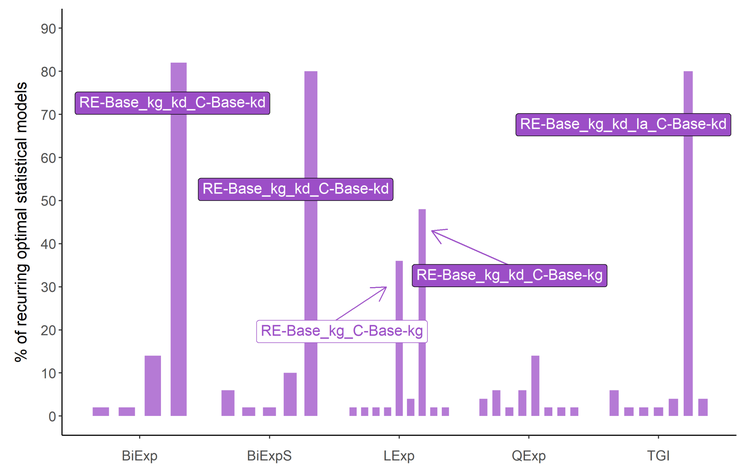
Results

Model Development

Only three empirical models (BiExp, BiExpS, and TGI) were able to stably converge to one base statistical model, i.e. they showed high reproducibility with the same optimal statistical model (in ≥50% of dataset splits) corresponding to the base model obtained with the full dataset.

However, the BiExpS model was excluded from the further analysis, because the estimate of parameter φ is approximately equal to 1 in all the scenarios, therefore BiExpS degenerates into BiExp model.

FIGURE 1. Statistical model reproducibility plot.



The boxes contain description of the base model' prevalence corresponding to the bars. RE - random effects, C - correlation between random effects. Filled boxes indicate frequently met statistical configurations that matched the ones obtained from the model qualification against the full, not split for training and validation, dataset.

The LExp has two comparable in magnitude peaks (met in 36% and 48% validation partitions) corresponding to statistical models differing in a single random effect on the tumor shrinkage parameter (k_g) pointing to the critical amount of information required to identify this parameter.

The QExp was excluded from the analysis, because of the lack of a consistent statistical model obtained against the training data for most validation partitions.

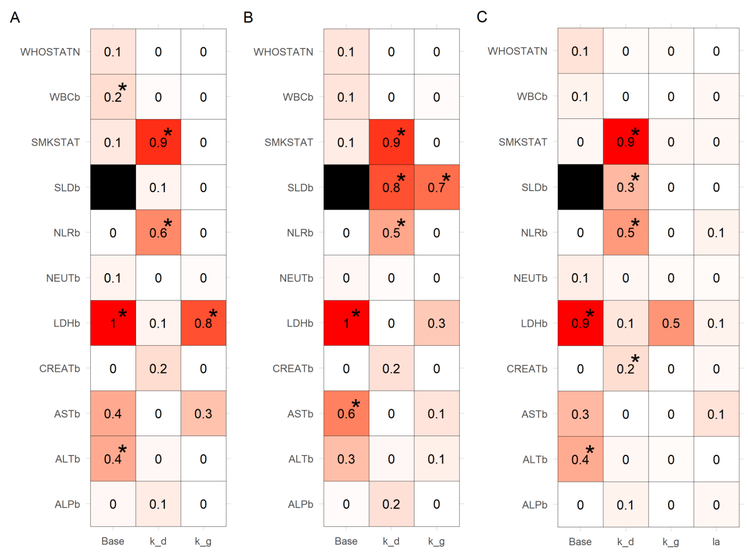


FIGURE 2. Heatmap of the frequency of inclusion of covariates in the final model via a stepwise covariate search algorithm in Monolix for (A) BiExp, (B) LExp, and (C) TGI. Base-SLDb combination (black cell) was excluded from covariate search. Covariates were also identified for the models tested against full data – these combinations are marked with asterisk.

Covariate-parameter relationship obtained in more than 50% validation tests were the same as for the models tested against the full dataset. The only exception was found for the baseline LDH concertation on k_g relation in the TGI model.

Therefore, the optimal covariate models were identified as:

BiExp: LDHb on $Base$; SMKSTAT, NLRb on k_d; LDHb on k_g;

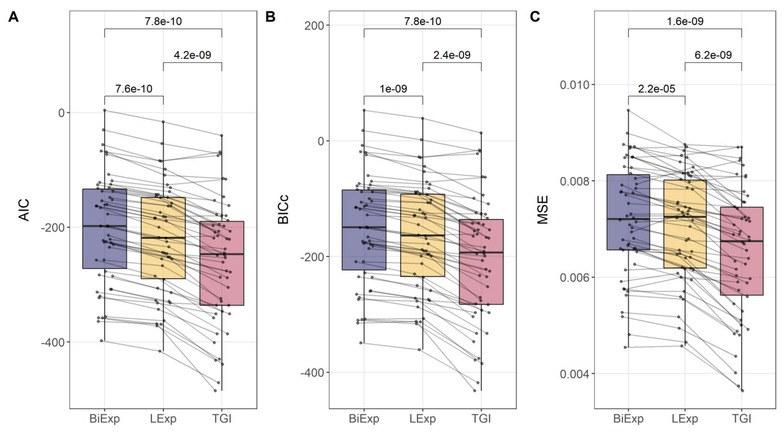
LExp: LDHb, ASTb on $Base$; SMKSTAT, SLDb, NLRb on k_d, SLDb on k_g;

TGI: LDHb on $Base$; SMKSTAT, NLRb on k_d;

Descriptive Power Assessment

Comparative analysis of final models descriptive ability revealed that the TGI model demonstrated superior descriptive ability, showing statistically significant differences in AIC, BICc and mean squared error (MSE) values compared to the BiExp and LExp. The LExp ranked second, indicating a clear hierarchy in model diagnostics performance.

FIGURE 3. Descriptive power assessment plot.

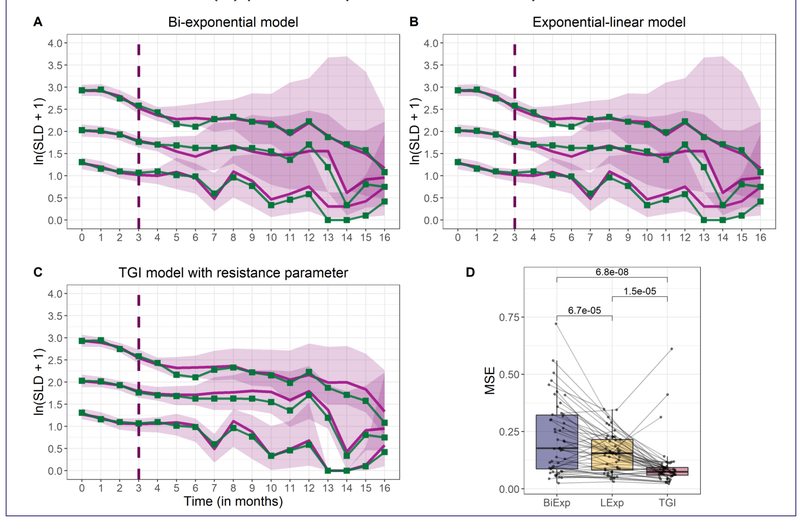


Predictive Power Assessment

Despite that the VPC plots for the validation data looked similarly for the three final models (Figure 4A-C), the TGI model significantly outperformed BiExp and LExp in predictive accuracy based on the calculated MSE values on validation datasets.

The boxplots also show that the variability of MSE decreases from the BiExp to the TGI model indicating higher stability of the predictive ability to original data splitting.

FIGURE 4 (A-C) Best run visual predictive check plots generated using validation datasets, (D) predictive power assessment plot.



Conclusions

- The presented work provides a methodological framework for empirical model optimization, offering a basis for more accurate predictions of tumor dynamics and model component choice for joint and sequential modeling framework.
- In this work we showed that there is a critical dependence of the final model on the available data.
- A simplified TGI model showed better performance than BiExp and LExp models against the investigated NSCLC patients data treated with EGFR tyrosine kinase inhibitor, erlotinib.

References

1. Zhudenkov K, et al. CPT Pharmacometrics Syst Pharmacol. 2022;11(4):425-437.
2. Baaz M, et al. CPT Pharmacometrics Syst Pharmacol. 2023;12(9):1227-1237.
3. Gonçalves A, et al. CPT Pharmacometrics Syst Pharmacol. 2024;13(1):68-78.
4. Bruno R, et al. *Clin Cancer Res*. 2023;29(6):1047-1055.
5. Eisenhauer EA, et al. *Eur J Cancer*. 2009;45(2):228-247.
6. https://data.projectdatasphere.org/projectdatasphere
7. Yin A, et al. CPT Pharmacometrics Syst Pharmacol. 2019;8(10):720-737.
8. Brum L. Guidance for Industry. Clin Pharmacol. Published online 2022.

PAGE, June 2024

Anna Mishina, anna.mishina@msdecisions.tech