Application of a Simulation-Based Diagnostics for Survival Evaluation in Joint Models

III-032

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

1 - M&S Decisions, Dubai, United Arab Emirates; 2 - Research Center of Model-Informed Drug Development, Sechenov First Moscow State Medical University, Moscow, Russia; 3 - Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences (INM RAS), Moscow, Russia

Introduction

- Joint models linking longitudinal biomarkers with time-to-event endpoints such as overall survival (OS) are increasingly used to support early efficacy assessment in oncology drug development.
- Model performance is typically evaluated using discrimination- and calibration-based metrics, such as time-dependent area under receiver operating curve (ROC AUC) and Brier Score. While these diagnostics assess the model's ability to distinguish between individuals at different risk levels and to predict survival probabilities, they are not sensitive to misspecification of the underlying survival distribution.
- This limitation is critical when models are used for tasks requiring accurate characterization of survival dynamics, such as extrapolation or treatment effect evaluation.

This work introduces simulation-based diagnostics designed to directly evaluate the adequacy of survival distribution predictions.

Methods

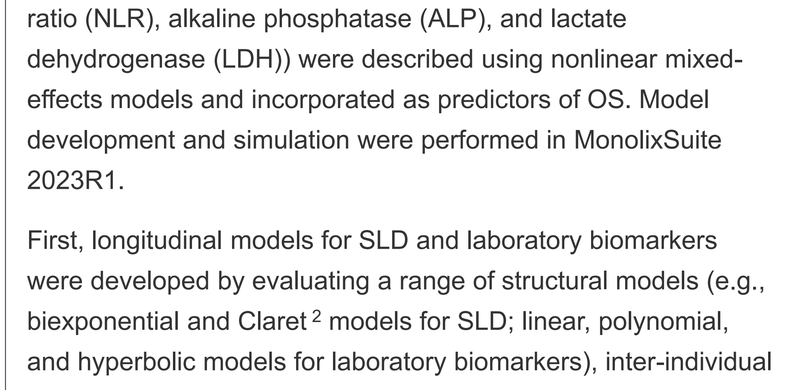
This study analyzed data from 381 patients with advanced non-small cell lung cancer (NSCLC) treated with Erlotinib (NCT00364351), sourced from Project DataSphere repository[1].

OS was analyzed using a joint modeling framework linking longitudinal biomarkers to time-to-event outcomes [2]. Tumor size, measured as sum of longest diameters of the target lesions (SLD), and selected laboratory biomarkers (neutrophil count (NEUT), white blood cell count (WBC), neutrophil-to-lymphocyte ratio (NLR), alkaline phosphatase (ALP), and lactate dehydrogenase (LDH)) were described using nonlinear mixed-effects models and incorporated as predictors of OS. Model development and simulation were performed in MonolixSuite 2023R1.

First, longitudinal models for SLD and laboratory biomarkers were developed by evaluating a range of structural models (e.g., biexponential and Claret [2] models for SLD; linear, polynomial, and hyperbolic models for laboratory biomarkers), inter-individual variability structures, and residual error models (additive, proportional, and combined). Model selection was based on Akaike Information Criterion (AIC) and parameter identifiability.

Second, joint models were constructed by linking longitudinal biomarker trajectories to OS, and different biomarker combinations were evaluated. Model performance was assessed using goodness-of-fit diagnostics and time-dependent discrimination and calibration metrics, including ROC AUC and Brier Score [2].

Third, two simulation-based diagnostic approaches were evaluated: (i) survival simulations based on empirical Bayes estimates (EBEs), and (ii) simulations conditional on observed longitudinal biomarker data:



Results

Final Joint Model Selection

- SLD biomarker was best described by Claret model, while the optimal structure for all the laboratory biomarkers was hyperbolic function.
- Among univariate joint models with identifiable parameters (relative standard error (RSE) < 51%), those incorporating hematological biomarkers demonstrated the highest prognostic performance.
- The joint model with neutrophil-to-lymphocyte ratio (NLR) as dynamic predictor demonstrated ROC AUC = 0.89 and BS = 0.14 at 12 months. In contrast, the model based solely on SLD showed inferior discrimination (AUC = 0.82, BS = 0.17), while models with alkaline phosphatase (ALP) or lactate dehydrogenase (LDH) performed the poorest performance (AUC ≤ 0.75, BS = 0.20).
- The NLR model was therefore selected as the backbone for multivariate extensions including SLD, ALP, and LDH. Adding SLD to the NLR model (NLR+SLD) improved discrimination over the univariate NLR model.
- However, further inclusion of ALP and LDH, either in three- or four-variate combinations, resulted in no improvement in discrimination performance.

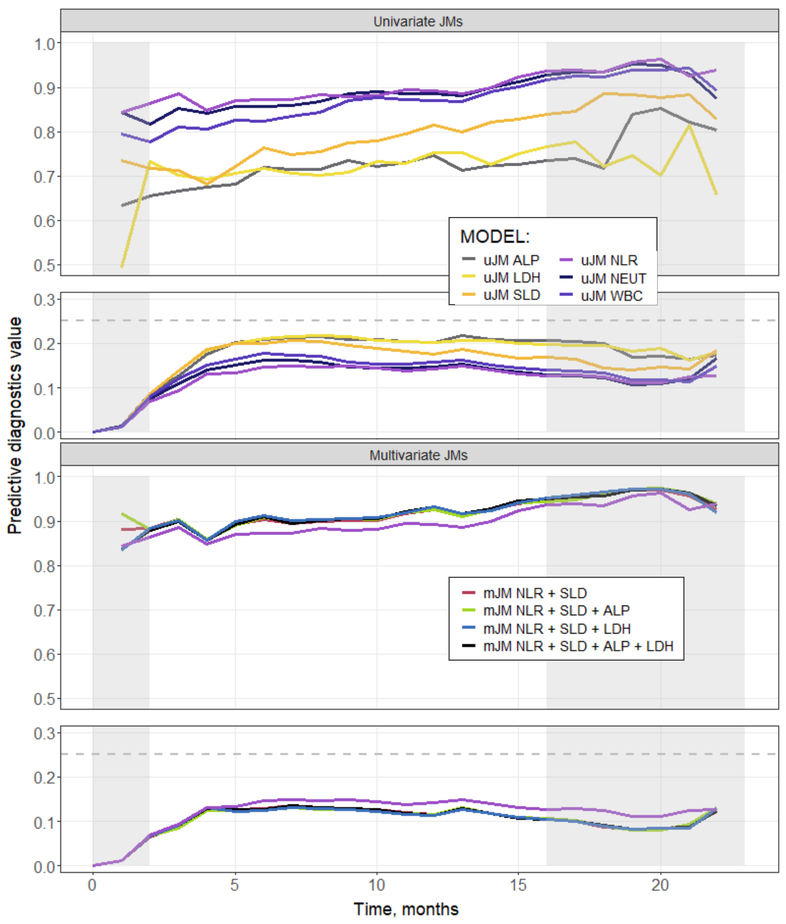


FIGURE 1. Discrimination performance of the developed joint models assessed via dynamic ROC AUC (top panel) and Brier score (bottom panel) over time, colored by model. Grey-shaded intervals indicate label event imbalance zones where ROC AUC may be not reliable.

- Final model was a joint longitudinal–survival model with:
  - Weibull baseline hazard function
  - two longitudinal biomarkers (SLD and NLR)
  - exponential link between biomarkers and hazard function
- High inter-individual variability in longitudinal parameters reflected the heterogeneous tumor and NLR dynamics of responders and non-responders within the advanced NSCLC population.
- Several parameters in the final joint model exhibited high shrinkage (Table 1). Therefore, the marginal distribution of individual parameters was biased towards population estimates [4].

Parameter Estimates			Magnitude of Inter-Individual Variability		
Parameter	Estimate	RSE%	CV%	RSE%	Shrinkage
Base	71.1	3.51	76.20	3.70	5.3
k_g	0.026	16.60	229.00	8.35	62.0
k_d	0.026	29.00	1,280.00	9.55	54.0
la	0.573		935.00	14.20	90.8
corr_k_d_Base	-0.34	22.70			
Constant RUV SLD	0.131	2.29			
hyp_c	2.14	9.53	173.00	7.27	38.6
hyp_b	1.33	2.44	47.10	4.25	10.2
hyp_l	2.82	3.54			
Constant RUV NLR	0.232	5.32			
Proportional RUV NLR	0.084	10.60	8.35		
shape	1.59	5.05			
scale	406	23.30			
beta_SLD	0.758	8.76			
beta_NLR	1.44	7.38			

TABLE 1. Parameter estimates for the final joint model. Summary of fixed effects, between-subject variability, and residual unexplained variability (RUV). RSE% are reported alongside each estimate.

Simulation-Based Diagnostics

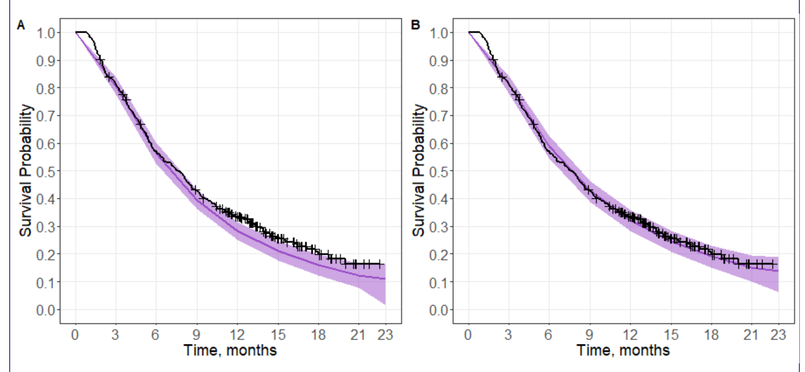


FIGURE 2. (A) EBE-based survival simulations and (B) Survival simulations conditional on observed longitudinal biomarker data. Simulated survival curves were smoothed, the violet solid line denotes the median simulated survival profile, and the violet shaded area represents the 95% PI across replicates;. The observed KM estimate is shown in black.

- EBE-based survival simulations were systematically biased relative to the observed Kaplan-Meier (KM) profile (Figure 2).
- This bias should therefore be interpreted with caution: it may reflect limitations of EBE-based simulation under high shrinkage rather than misspecification of the longitudinal or survival submodels.
- Sampling from the conditional distribution allows to evaluate unbiased population distribution of individual parameters [4].
- Because survival simulations propagate individual parameters through nonlinear, time-integrated hazard functions, shrinkage-induced distortion of the EBE-based latent trajectories can translate into biased KM profiles.
- Conditional survival simulations demonstrated close agreement between predicted and observed population survival function (Figure 3).
- This supports the use of the final joint model for dynamic survival prediction, given observed longitudinal biomarker data.
- However, this does not by itself validate the model for unconditional population-level trial simulation, where individual parameters are generated de novo from the model-implied inter-individual variability distribution (Ω).

Conclusions

- The proposed conditional simulation-based diagnostic provides an additional tool to assess whether the joint model produces realistic survival predictions for the biomarker trajectories observed in the study population.
- In the presence of substantial shrinkage, conditional simulations are preferable to EBE-based simulations for evaluating survival calibration conditional on observed longitudinal data.
- Further evaluation using Ω-sampled individual parameters is required to assess the model's ability to reproduce marginal survival outcomes in replicated population-level trial simulations.

References

1. https://data.projectdatasphere.org/projectdatasphere
2. Rizopoulos, D. Joint Models for Longitudinal and Time-to-Event Data.
3. Claret, L.; Girard, P.; Hoff, P. M.; Van Cutsem, E.; Zuideveld, K. P.; Jorga, K.; Fagerberg, J.; Bruno, R. Model-Based Prediction of Phase III Overall Survival in Colorectal Cancer on the Basis of Phase II Tumor Dynamics. JCO 2009, 27 (25), 4103–4108. https://doi.org/10.1200/JCO.2008.21.0807
4. Lavielle, M.; Ribba, B. Enhanced Method for Diagnosing Pharmacometric Models: Random Sampling from Conditional Distributions. Pharm Res 2016, 33 (12), 2979–2988. https://doi.org/10.1007/s11095-016-2020-3.

Poster III-032, 34th PAGE meeting, Dubrovnik, Croatia

Anna Mishina
anna.mishina@msdecisions.tech
LinkedIn
Website