

TUTORIAL

A workflow for the joint modeling of longitudinal and event data in the development of therapeutics: Tools, statistical methods, and diagnostics

Kirill Zhudenkov¹ | Sergey Gavrilov^{1,2} | Alina Sofronova¹ | Oleg Stepanov¹ |
Nataliya Kudryashova¹ | Gabriel Helmlinger³ | Kirill Peskov^{1,4}

¹M&S Decisions LLC, Moscow, Russia

²The faculty of Computational Mathematics and Cybernetics, Lomonosov MSU, Moscow, Russia

³Clinical Pharmacology & Toxicology, Obsidian Therapeutics, Cambridge, Massachusetts, USA

⁴Research Center of Model-Informed Drug Development, Sechenov First Moscow State Medical University, Moscow, Russia

Correspondence

Kirill Zhudenkov, M&S Decisions, 5, Naryshkinskaya al. Moscow 125167, Russia.

Email: kirill.zhudenkov@msdecisions.ru

Funding information

This work was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital biodesign and personalized healthcare" no. 075-15-2020-926.

Abstract

Clinical trials investigate treatment endpoints that usually include measurements of pharmacodynamic and efficacy biomarkers in early-phase studies and patient-reported outcomes as well as event risks or rates in late-phase studies. In recent years, a systematic trend in clinical trial data analytics and modeling has been observed, where retrospective data are integrated into a quantitative framework to prospectively support analyses of interim data and design of ongoing and future studies of novel therapeutics. Joint modeling is an advanced statistical methodology that allows for the investigation of clinical trial outcomes by quantifying the association between baseline and/or longitudinal biomarkers and event risk. Using an exemplar data set from non-small cell lung cancer studies, we propose and test a workflow for joint modeling. It allows a modeling scientist to comprehensively explore the data, build survival models, investigate goodness-of-fit, and subsequently perform outcome predictions using interim biomarker data from an ongoing study. The workflow illustrates a full process, from data exploration to predictive simulations, for selected multivariate linear and nonlinear mixed-effects models and software tools in an integrative and exhaustive manner.

INTRODUCTION

Clinical trials are a gold standard in the development of novel therapeutics. Through clinical studies, information and data are generated and may be integrated on a continuous basis, thereby improving our understanding of disease pathology and progression, pharmacologic

intervention, trial design, and, ultimately, personalized medicine.^{1,2}

Clinical outcomes may be assessed based on a number of criteria.³ In earlier stages of development, longitudinal biomarkers may be measured according to a defined schedule, allowing for the investigation of patient response trends and to determine whether a biomarker reaches a

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 M&S Decisions LLC. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

target value. In later phases of clinical studies and often in addition to longitudinal biomarkers, it is common to assess endpoints such as time-to-event data or event rates. Depending on the disease or indication, outcomes may encompass overall survival (OS) or progression-free survival (in oncology), exacerbation risk and/or rates (e.g., chronic obstructive pulmonary disease [COPD], asthma), relapse rates (e.g., multiple sclerosis), myocardial infarction, stroke, or other cardiovascular events (e.g., cardiovascular diseases, chronic kidney disease [CKD]), and so on.³

Intuitively, one may expect that baseline and/or longitudinal biomarkers are correlated to some degree with observed event outcomes. For example, the forced expiratory volume in 1 s is associated with exacerbation risk in COPD⁴; the sum of the longest diameters of target lesions (SLD) is associated with risk of death in non-small cell lung cancer (NSCLC).⁵ The association between biomarker and time-to-event data has also been identified and studied in multiple analyses of various diseases, including heart failure, CKD, diabetes, Alzheimer's disease, infectious diseases, colorectal and prostate cancers, and so on.^{6–13} These analyses have investigated the association of biomarkers and time-to-event data via joint models (JMs).¹⁴ JMs represent an extended class of parametric survival models, which are based on proportional hazards models and are able to integrate both baseline values and dynamics of longitudinal covariates. JMs may ultimately be used to address essential questions in the assessment and prediction of clinical trial outcomes, including investigation and prediction of longitudinal biomarker trends, biomarker association with an event risk, individual survival predictions, and unbiased estimation for both biomarker and time-to-event outcomes.^{3,6,14}

In this tutorial, we focus on technical aspects of a modeling workflow for a common class of JMs that incorporate noninformative right censoring and proportional hazards survival submodels. Right censoring is a specific property of time-to-event data and is frequently encountered in clinical trials. An event of interest does not necessarily occur for all subjects during the trial observation period, and the time to that event, if an event occurs at all, will vary among trial subjects. A censoring event occurs when a subject leaves the study for a specific reason. Depending on the dropout nature, censoring could be classified as informative or noninformative.¹⁴ Informative censoring assumes that the censoring event is directly related to the subject's disease progression, whereas noninformative censoring assumes the censoring is independent from the subject's condition and therefore from the event of interest. The noninformative hypothesis is commonly used for survival modeling in clinical trials.^{6,10,14}

The JMs described in this tutorial may be developed using various software tools integrating linear and non-linear mixed-effects (LME and NLME, respectively) approaches to handle longitudinal data.

This modeling workflow seamlessly integrates multiple steps, including data exploration, survival model selection, and qualification (goodness-of-fit and parameter interpretation) using training data and subsequent model validation against interim external data accumulated up to a certain timepoint (Figure 1). Such a workflow can be applied by a data analyst or modeling scientist who, upon collecting retrospective information and data from completed clinical trials, aims at predicting outcomes of another clinical study, for which only interim data are available.

Using particular NSCLC data sets,^{15,16} we implemented and tested the JM workflow to investigate the association between selected biomarkers and OS and to quantify the incremental benefit that is gained using longitudinal biomarker data versus baseline biomarker only.

STRUCTURE OF JMS

Biomarker and event data may be described using various regression models.^{17,18} LME and NLME models are typically developed to characterize longitudinal biomarkers:

$$\begin{cases} y_i(t) = X_i(t)\beta + Z_i(t)b_i + \varepsilon_i(t), \\ b_i \sim N(0, D), \\ \varepsilon_i \sim N(0, \sigma^2 I_{n_i}), \end{cases} \quad (1)$$

Equation (1) represents a linear model, where X_i and Z_i are known design matrices for, respectively, fixed-effects regression coefficients β and random-effects coefficients b_i , with i representing the patient index, where I_{n_i} denotes the n_i -dimensional identity matrix.¹⁴ Random effects are assumed to follow a multivariate normal distribution (because multiple biomarkers and model parameters may be considered) with a mean vector of 0 and a variance-covariance matrix D and are also assumed to be independent from the error terms ε_i , that is, $\text{cov}(b_i, \varepsilon_i) = 0$. Further details on univariate and multivariate model structures for longitudinal biomarkers and their interpretation may be taken from the literature.^{14,19}

Proper selection of model structure and subsequent qualification of LME and NLME models may provide the modeler with biomarker trends in the population under study as well as information on random effects and residual error as typically performed in pharmacometric research.²⁰ Similarly to previous analyses, one may use predetermined NLME models defined in the form of

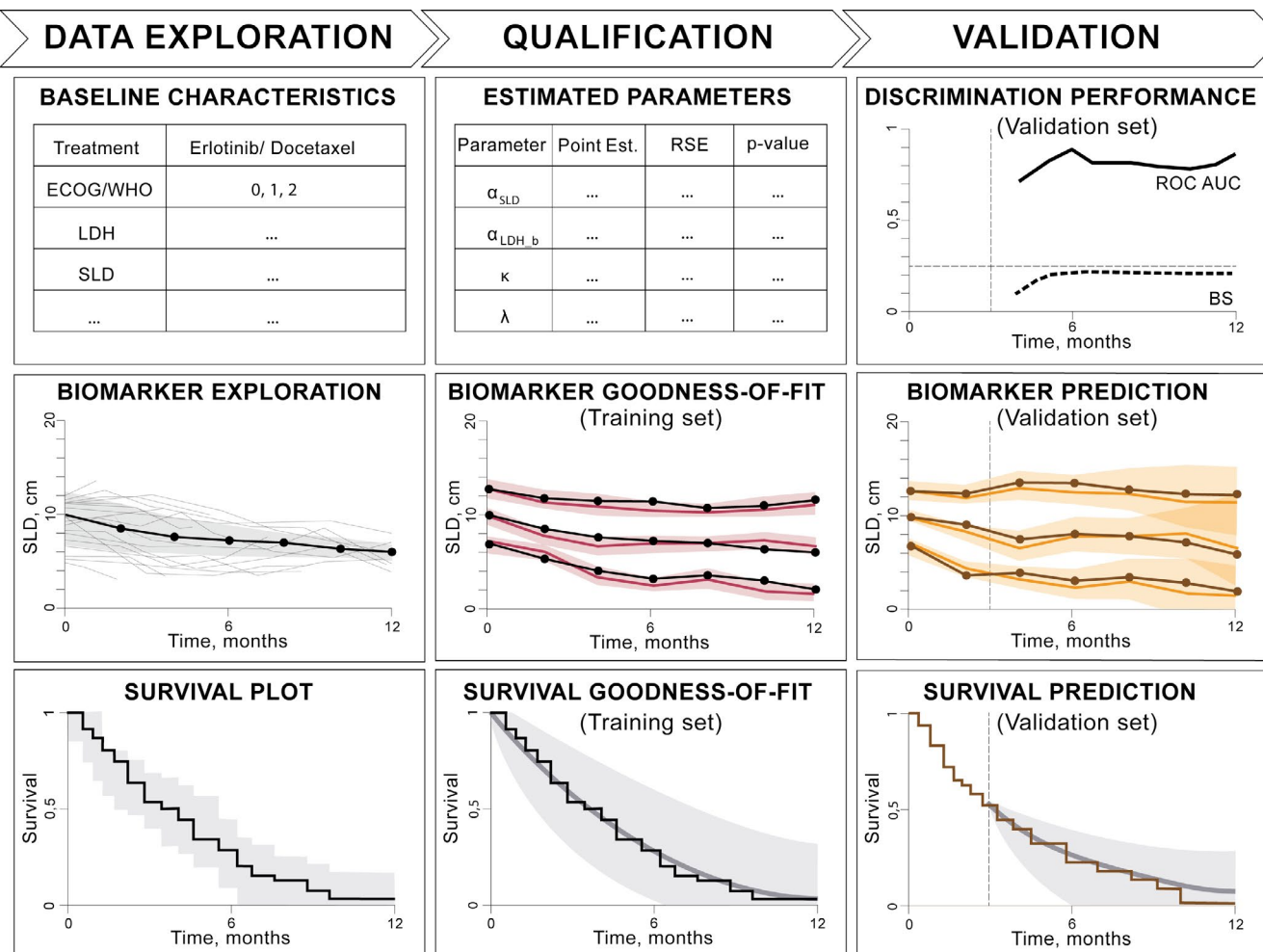


FIGURE 1 Schematic representation illustrating the steps required for a joint modeling analysis. Left to right: data exploration involves plotting biomarker and survival data as well as study baseline characteristics. Qualification of survival models is performed against a training data set to obtain model parameters, visual predictive checks for biomarkers, and survival goodness-of-fit plots. Validation of a joint model is completed by using the previously estimated model parameters to perform a survival discrimination analysis by means of ROC-AUC and BS as well as simulate biomarker trends and make survival predictions against interim validation data. BS, Brier score; ECOG, Eastern Cooperative Oncology Group score; LDH, lactate dehydrogenase; ROC-AUC, area under the receiver operating characteristic curve; SLD, sum of the longest diameters of target lesions; WHO, World Health Organization score.

explicit/empirical functions or using ordinary differential equations (ODEs).¹⁷ However, for many biomarkers, such information may not be available, and additional research would be required to identify an adequate model.

Time-to-event data may be analyzed using survival models, which are defined by a hazard function¹⁴:

$$h_i(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T_i^* < t + dt | T_i^* \geq t)}{dt}, \quad t > 0 \quad (2)$$

Equation (2) describes the instantaneous risk of an event in the time interval $[t, t + dt]$, provided survival up to time t . Given this definition, $h_i(t)$ is also referred to as the instantaneous risk function. T_i^* denotes the true time of events for the patient of interest. The hazard function

defined previously can be used to derive individual survival probabilities according to the following equation¹⁴:

$$S_i(t) = \exp \{ -H_i(t) \} = \exp \left\{ - \int_0^t h_i(s) ds \right\} \quad (3)$$

The survival function (whether it represents actual survival, or progression, exacerbations, relapses, etc., depending on the disease and indication under study) can be expressed by means of a cumulative hazard function $H_i(t)$ that describes the accumulated risk up to time t . Function $H_i(t)$ can also be interpreted as the expected number of events to be observed up to time t . Individual survival functions $S_i(t)$ can be further used to estimate the survival of the overall population and may be compared with a

nonparametric survival estimator by means of a Kaplan-Meier (KM) curve.

JMs may be used to associate baseline (w_i) and longitudinal covariates (represented by the individual biomarker time course: $m_i(t) = X_i(t)\beta + Z_i(t)b_i$ from Equation (1) with event risk from Equation (2). For such proportional hazards models, $h_i(t)$ may be formulated as follows:

$$h_i(t) = h_0(t) * \exp \{ \alpha_b w_i + \alpha m_i(t) \}, \quad (4)$$

In Equation (4), α_b represents the association parameters for the baseline biomarkers (as typically implemented in Cox models), and α is the association parameter for the longitudinal biomarker. Equation (4) is formulated for one baseline and one longitudinal biomarker. For multivariate models, each of the considered biomarkers should have α_b or α assigned depending on a biomarker type.

If biomarkers are considered at baseline only, then Cox proportional hazards models can be used and $h_0(t)$ is not specified because the Cox model parameter optimization is handled using partial log-likelihood function. However, in JMs with longitudinal biomarkers, $h_0(t)$ should be specified in either way (i.e., piecewise constant function, splines, a Weibull distribution).^{14,21} Parameters of a JM can be identified using the principle of maximum likelihood based on the following formulation for the logarithmic joint conditional likelihood function^{14,22}:

$$l_i(\theta) = \log \int \left\{ \prod_j^{n_i} p(y_{ij}|b_i; \theta) \right\} \{ h(T_i|b_i; \theta)^{\delta_i} S_i(T_i|b_i; \theta) \} p(b_i; \theta) db_i,$$

where:

$$S_i(t|b_i; \theta) = \exp \left(- \int_0^t h_0(s; \theta) \exp \{ \alpha_b w_i + \alpha m_i(s) \} ds \right) \quad (5)$$

In Equation (5), δ_i is an event indicator (i.e., assumes a value of 1 for an event and of 0 for censoring from the study), T_i is the observed event time or censoring time ($T_i = \min(T_i^*, C_i)$, where C_i is the censoring time and θ denotes the full vector of model parameters). Averaged $S_i(t)$ represents the marginal survival for the entire patient cohort and can be compared with KM estimates of OS.¹⁴

Other approaches that incorporate a sequential analysis of biomarkers and time-to-event data (including extended Cox models that handle time-varying biomarkers) have also been investigated in the literature. However, such analyses may require methodological improvements to address the potential bias presented in parameter estimates.^{14,21,23}

OVERVIEW OF SELECTED TOOLS FOR JOINT MODELING

There is a wide variety of statistical software tools available for JM development and assessment—see Table 1 for a nonexhaustive list. Each JM software comes with specific sets of options as well as data set format requirements. For example, if a simple univariate analysis is targeted in the *JM* package²⁴ in R, then only one longitudinal biomarker may be considered, whereas the other available biomarkers can only be accounted at baseline levels. Beyond this limitation, the *JM* package provides a comprehensive set of options, such as predefined functions for model qualification/diagnostics (embedded in the JM object in the R environment) and validation/simulation (available via *predict()* and *survfitJM()* functions), and it is computationally fast.

If models with multiple longitudinal biomarkers are considered, the *JMbayes* and *rstanarm* software packages,^{25,26} which provide Bayesian inference for biomarker characterization, can be used. However, in these multivariate JM packages, the advanced diagnostics for longitudinal biomarkers are less convenient to handle, and the embedded features for biomarker/survival simulations have limited functionality. Also, these packages require similar measurement times for all considered longitudinal biomarkers (often it is not the case in actual clinical data sets). Moreover, all of these packages (*JM*, *JMbayes* and *rstanarm*) reach limitations when it comes to the description of biomarker dynamics because only linear functions for biomarker description may be used. Thus, if the dynamics of biomarkers need to be investigated with more mechanistic models, these R-based LME packages may not represent ideal choices. Nevertheless, with some data

TABLE 1 Joint modeling tools, their features, diagnostics, and simulation capabilities

	<i>JM</i> package	<i>JMbayes</i> package	<i>rstanarm</i> package	NONMEM	Monolix	Stan
LME models	Yes	Yes	Yes	Yes	Yes	Yes
NLME models	No	No	No	Yes	Yes	Yes
Multivariate models	No	Yes	Yes	Yes	Yes	Yes
Biomarker and survival predictions	Available	Limited	Limited	Available	Available	Available

Abbreviations: LME, linear mixed effects; NLME, nonlinear mixed effects.

sets, linear models, for example, based on splines, may adequately characterize biomarker dynamics.⁴

NLME models have become a standard in quantitative clinical pharmacology for longitudinal biomarker analysis.²⁷ These models represent a natural choice to describe these biomarkers and to associate their dynamics with event risk in a JM framework. For such advanced nonlinear JMs, the widely known NONMEM and Monolix pharmacometric software tools may be used.^{28–30} NLME JMs can also be implemented in Stan software, with even more advanced options including user-defined likelihood function formulation.^{5,6,31,32}

In this tutorial, and for illustrative purposes, we implemented a JM workflow using the following set of modeling tools: *JM* v1.4–8 and *JMbayes* v0.8–85 packages for LME JMs, and Monolix 2020R1 for NLME JMs. All R-based packages were tested in the R 4.0.2 environment.³³

CASE STUDY: DATA DESCRIPTION AND EXPLORATION

Data from subsets of control arms from two NSCLC clinical studies, NCT00312377¹⁶ and NCT00364351,¹⁵ were taken from the Project Data Sphere repository.³⁴ In both subsets, patient data had been deidentified and did not include data from Chinese patients. Biomarker and OS data from the NCT00312377 study were used for model qualification (training data set); data from the NCT00364351 study were used for external model validation (validation data set), according to the workflow illustrated in Figure 1.

The treatments administered to these subjects with advanced NSCLC were as follows: chemotherapy (docetaxel) in the training data set and targeted therapy (erlotinib) in the validation data set. To build a JM framework, we focused on prognostic biomarkers, which may potentially provide information on survival outcomes regardless of treatment type. Interestingly, patient baseline characteristics across the two selected data sets matched rather well, including the percentage of patients with mutations in the epidermal growth factor receptor—see Table S1.

For simplicity, we limited the number of investigated biomarkers to two: tumor size and levels of lactate dehydrogenase (LDH).^{35,36} Tumor size was represented by SLD, measured according to the Response Evaluation Criteria in Solid Tumors 1.1.³⁷ LDH was selected as a biomarker because it is a measure linked to the magnitude of tumor metabolic activities. Data for both selected biomarkers were available at baseline as well as longitudinal measurements, with a sampling schedule defined in Table S1. Exploratory plots of the selected biomarkers as well as OS data are shown in Figure 2.

When examining these data sets across the two chosen clinical studies, likeness in trends can be observed for SLD, whereas differences in both LDH and OS were noted. Given the similarity in patients' baseline characteristics across the two studies, the apparently different patterns in longitudinal LDH and OS may be caused by the administered therapies. The observed trends in biomarker dynamics may also be dependent on the study design, measurement schedule, and patient censoring, whereas at longer times, only surviving patients with smaller tumor sizes and lower LDH levels typically remain in the study.

With the exploratory plots at hand (Figure 2), we next describe the subsequent steps in the data processing and JM model qualification and validation illustrated schematically in Figure 1. Specifically, there we represent the analysis and diagnostics for the association of baseline and longitudinal biomarkers with OS. For more details on the JM workflow and its description, see the Supplementary Materials, Figure S1.

DATA SET PREPARATION

A standard vertical data set can be prepared for Monolix^{38–40} as described in the Supplementary Materials, Section 1.1. The data frame may be saved as a comma-separated values (CSV) file to develop and test models via the graphical user interface or may be kept as an object in the R environment, if *lixoftConnectors/RsSimulx* packages are used to control Monolix.⁴¹

The R-based LME tools used in this tutorial require a data set format that is common for conventional analyses of longitudinal data (in *nlme*⁴² or *lme4*⁴³ packages) and time-to-event data (in *survival*⁴⁴ package) in R. All data can be incorporated into a single data frame. We provide suggestions on data set structuring for *JM/JMbayes* in the Supplementary Materials, Sections 1.2 and 1.3.

An actual view of the formatted data, ready for further analyses in *JM/JMbayes* and Monolix, is provided in the Supplementary Materials, Figure S2.

Different biomarker data transformations may be considered for optimal model convergence (e.g., \log_2 , \log_{10} , square root).¹⁴ In the present data set and because of the high dynamic ranges observed (Figure 2), SLD and LDH data were log-transformed into $\ln(\text{SLD}+1)$ and $\ln(\text{LDH}+1)$.

SURVIVAL MODEL BUILDING

In survival modeling, the point estimates of parameters obtained in the model outputs should be studied. The size of the parameter association with OS can be assessed using the hazard ratio (HR) value, which is calculated as $\exp(\alpha_b)$ from Equation (4) for baseline or as $\exp(\alpha)$ for longitudinal

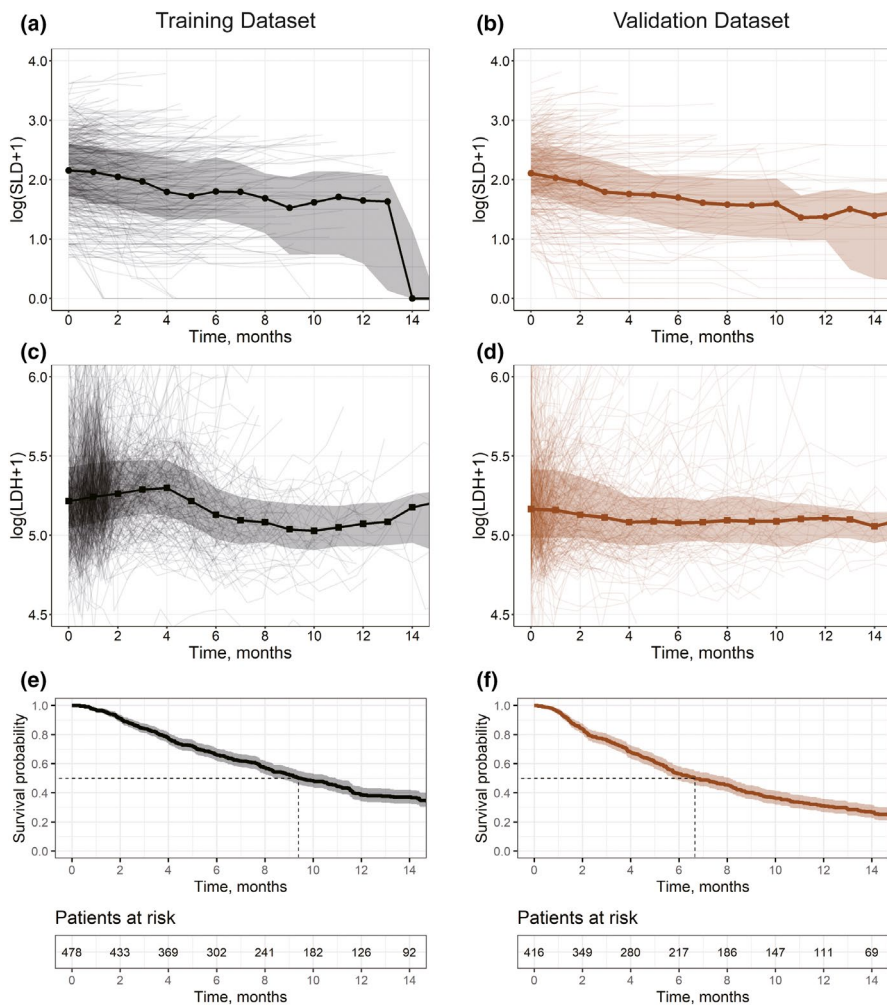


FIGURE 2 Exploration plots for training (black, left) and validation (brown, right) data sets. (a–d) Spaghetti plots for log-transformed biomarkers showing mean values (solid lines; circles, SLD; squares, LDH) and interquartile range (shaded area). (e–f) Survival plots showing experimental Kaplan-Meier curves (solid lines) and 95% confidence interval (shaded area). LDH, lactate dehydrogenase; SLD, sum of the longest diameters of target lesions

biomarkers. A covariate increase by one unit would result in an instantaneous hazard change by a corresponding HR value.¹⁴ Thus, in addition to providing statistical significance for a biomarker (a p value of <0.05 should be taken¹⁴), HR provides the inference for clinical significance similarly to a conventional pharmacometric analysis.⁴⁵

In some cases, a particular pattern in biomarker dynamics may not have a profound effect on time-to-event outcomes, although these statistically significant biomarkers may be associated with an improved quality of life, patient-reported outcomes, or other aspects related to disease burden.⁴ There may also be situations when biomarker measurement noise is high and/or data sampling is too sparse (e.g., biomarkers of interest are measured only at baseline and at the end of a study), so that the actual trends cannot be properly captured, and model qualification cannot be performed adequately.

Cox proportional hazards models for baseline biomarkers

An initial and simple step in the analysis of association between biomarkers and a time-to-event endpoint may

include selection and testing of relevant baseline biomarkers via semiparametric Cox proportional hazards models that are conventionally used in clinical study analysis and offer fast parameter optimization.⁴⁶ For further survival analysis involving JMs, these initially identified statistically significant biomarkers may be investigated longitudinally by means of fully parametric JMs.

Cox proportional hazards models may be developed and tested using *survival* package in R via the *coxph()* function⁴⁴ and with the following syntax:

```
COX <- coxph(Surv(EVENT_TIME,EVENT) ~ SLDb + LDHb, data = jmdata.id, x = TRUE)
```

As seen from the model code, we investigated the importance of two baseline biomarker candidates, SLD and LDH. The optimal model should incorporate both statistical significance of the tested biomarkers (via a p value assessment for the association parameters) and the lowest value for a chosen penalized likelihood statistical criterion (i.e., Akaike information criterion [AIC]), which describes the relative goodness-of-fit of the tested models against the training data set. We

considered the following three survival models: univariate Cox SLD, univariate Cox LDH, and multivariate Cox SLD+LDH, represented in the aforementioned code and hereafter referred to as “COX.” The outputs of these tests (Table S2) suggest that the multivariate COX model should be selected for further research.

In terms of parameter estimates in the survival models developed here, it should be noted that the size of the biomarker effects (i.e., the estimates of the association parameters) may change when additional baseline or longitudinal covariates are introduced into the model. Therefore, the summary effect on event risk may get spread over a broad range of the chosen statistically significant biomarkers. In particular, it can be noticed from Table S2 that in univariate Cox models, the value of SLD and LDH association parameters have smaller magnitude than in the multivariate model. In further data analysis, one may also include and test additional biomarkers for their association with OS, for example, smoking status, Eastern Cooperative Oncology Group (ECOG) score, and so on.⁵

Following the analysis of baseline SLD and LDH, we next investigate the association of longitudinal biomarker trends with OS in the training data set (Figure 1).

Linear JMs

As mentioned previously, various tools may be used for JM development. An easy way to investigate longitudinal biomarkers and event risk is to develop linear univariate or multivariate JMs using the *JM/JMbayes* packages in R.¹⁴ The codes for the tested LME JMs are available in the Supplementary Materials, Section 2.1.

In these models, the dynamics of longitudinal biomarkers are described using natural spline functions.¹⁴ The knot position of the splines may be updated or additional covariates may be introduced to further optimize the description of biomarkers, depending on the available measurement frequency.

The survival submodels for univariate JMs¹⁴ (JM SLD and JM LDH for longitudinal SLD and longitudinal LDH, respectively) were developed in the *JM* package using a Weibull formulation for the baseline hazard. To reach the optimal model convergence for a multivariate SLD and LDH model (JM SLD+LDH), the model was built in the *JMbayes* package using a spline representation of the baseline hazard.

Nonlinear JMs

Advanced JMs can be developed and tested with NLME models for biomarkers. A full representative Monolix

model code is provided in the Supplementary Materials, Section 2.2. In this analysis, we considered biexponential functions to describe SLD and LDH dynamics^{5,17} using the following ODE formulation:

$$\begin{aligned} \frac{dDegr}{dt} &= -d * Degr \\ \frac{dGrowth}{dt} &= g * Growth \\ Biom &= Biom_b(Degr + Growth - 1) \\ \ln Biom &= \ln(Biom + 1), \end{aligned} \quad (6)$$

where d and g are parameters representing rate constants of biomarker dynamics, and $Biom_b$ is the estimated baseline value for a biomarker. Equation (6) represents a longitudinal submodel for either of the chosen biomarkers: SLD or LDH. These models can also be formulated by means of explicit nonlinear functions. A set of nonlinear models similar to the one for LME tools has been built in Monolix (JM SLD, JM LDH, and JM SLD+LDH). Univariate models may be generated by removing relevant longitudinal biomarker sections (either for SLD or LDH) from the model code and considering this biomarker at baseline only.⁴⁷

For these NLME JMs, the hazard function embeds the Weibull formulation of the baseline hazard, and it can be represented as follows:

$$h_i(t) = \frac{k}{\lambda} * \left(\frac{t}{\lambda}\right)^{k-1} * e^{(\alpha_j * \log Biom_{ij} + BaseCov)}, \quad (7)$$

where parameters k and λ define the shape of the Weibull baseline hazard function, and α_j is the association parameter for the j th longitudinal biomarker (either SLD or LDH). There are no random effects added to k , λ , and α_j .¹⁴ In Monolix survival models, a special parameter *BaseCov* is used to introduce baseline covariates (its fixed-effect value should be set to 0, and random effects should be turned off).^{29,47} The baseline covariates of interest should be associated with this *BaseCov* parameter; thus, the respective association coefficients will reflect the values of α_b in Equation (4). Finally, in univariate JMs, if SLD is considered as longitudinal, then LDH is taken at baseline only and vice versa.

Monolix incorporates a set of embedded tools to check for optimal model convergence. Typically, for successful model convergence and usability toward advanced diagnostics and predictions, one should pay attention to the available recommendations.^{30,48–51} More details are provided in the Supplementary Materials, Section 2.3.

MODEL DIAGNOSTICS

Detailed summaries of univariate and multivariate JMs for the selected LME and NLME tools are provided in the

Supplementary Materials, Table S3. The association parameters for baseline and longitudinal biomarkers and OS are provided in Table 2. These results suggest a similarity in biomarker association with OS in the training data set across the same JMs developed with different tools.

Advanced graphical diagnostics of the developed JMs include various plots to efficiently assess the goodness-of-fit for the biomarker and time-to-event data description (Figure 3). These graphical diagnostics can be generated in R environment for both LME tools and Monolix in a similar manner—see the Supplementary Materials, Section 3.

Individual predictions of biomarker trends and $S_i(t)$ in Monolix were retrieved by performing a sampling from the conditional distribution of individual biomarker sub-model parameters using a Markov Chain Monte Carlo (MCMC) procedure (50 simulated parameters per patient) that may be controlled in the Monolix graphical user interface or using the *lixoftConnectors/RsSimulx* tools in R.^{14,52}

Visual predictive checks (VPCs) provide further information on how well a model describes or predicts longitudinal data. The observed data were represented

with medians and 10% and 90% quantiles of observations using a moving average of a fixed length (e.g., 30 days). Biomarker trend predictions were obtained only for time-points corresponding to actual observations. Similarly to the observed data, the predicted values were drawn using a moving average: solid lines represented 10%, 50%, and 90% quantiles of averaged individual predictions. Prediction intervals for each quantile were computed with a 95% confidence interval. VPCs were plotted against the training data, thus describing the goodness of fit.

Patient survival plots (middle column, bottom graph in Figure 1) represent an important step in the model qualification assessment. They illustrate the alignment of the survival function from a qualified JM and actual KM estimates for the training data. Per timepoint, the mean value of the sampled individual survival probability (see Equation 5) is further averaged across all subjects to derive the predicted marginal survival. The range over the predicted marginal survival represents the 25th and 75th percentiles of the aggregated individual survival predictions. For JMs developed in Monolix, the individual survival function $S_i(t)$ was manually

TABLE 2 Association parameters for all tested joint models

Model type	R-based tools			Monolix tools		
	JM SLD	JM LDH	JM SLD+LDH	JM SLD	JM LDH	JM SLD+LDH
α_{SLD}	0.77 (13.3)	0.53 (20.5)	0.7778 (12.9)	0.85 (7.41)	0.54 (6.06)	0.75 (7.35)
α_{LDH}	0.44 (29.7)	0.75 (14.8)	0.4581 (22.1)	0.38 (13.0)	0.78 (3.57)	0.60 (7.85)
k	1.31 ^a	1.34 ^a	N/A ^b	1.28 (2.3)	1.26 (2.2)	1.26 (1.82)
λ	6520 ^a	14,300 ^a	N/A ^b	8380 (15.9)	28,100 (22.6)	19,300 (16.4)

Note: Values in bold font indicates the association parameters for longitudinal biomarkers tested in the models. Values are provided as point estimates (%RSE). Abbreviations: JM, joint model; LDH, lactate dehydrogenase; SLD, sum of the longest diameters of target lesions.

^aDerived from JM summary (see Supplementary Materials, Table S3).

^bBecause Weibull baseline hazard parametrization is not available (N/A) in *JMbayes*, splines were used instead.

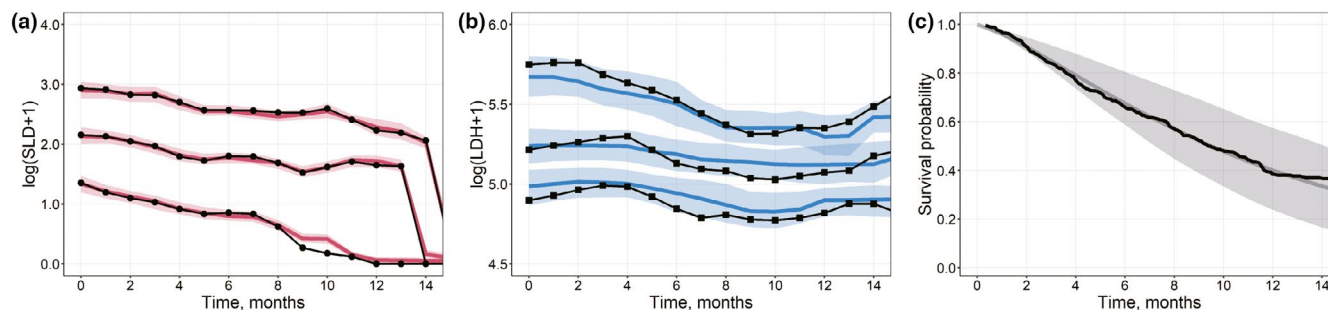


FIGURE 3 (a, b) Visual predictive check plots for the training data set (Qualification Step). Model results obtained using a multivariate nonlinear biexponential joint model in Monolix. Solid black lines represent aggregated log-transformed biomarker data from the training data set (circles, SLD; squares, LDH). Solid colored lines represent 10%, 50%, and 90% quantiles of averaged individual predictions (red, SLD; blue, LDH). Prediction intervals for each quantile are computed with a 95% confidence interval (shaded area). (c) Survival plot shows experimental Kaplan-Meier curve (solid black) from the training data set and the mean model prediction (solid gray) with the interquartile range (shaded area). LDH, lactate dehydrogenase; SLD, sum of the longest diameters of target lesions

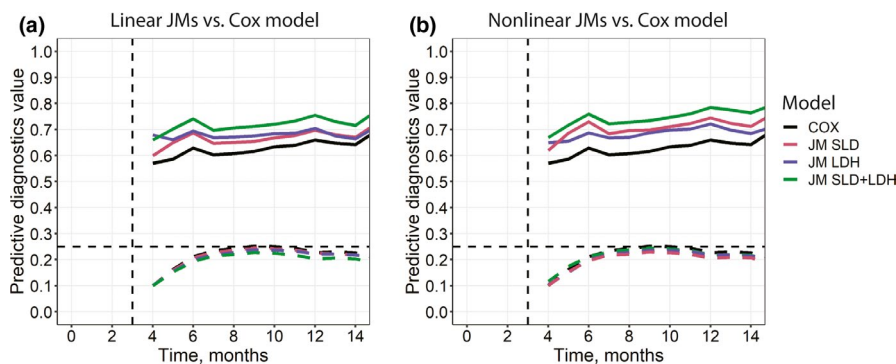


FIGURE 4 Area under the receiver operating characteristic curve (solid line) and Brier score (dashed line) diagnostics for the interim validation data set based on the selected survival models. (a) Linear models in *JM/JMbayes* packages. (b) Nonlinear models in *Monolix*. Both plots feature the same COX model, which was qualified using the *coxph()* function in R (see the Supplementary Materials, Section 4.2, Table S2 and “Cox Proportional Hazards Models for Baseline Biomarkers” section in the main text). COX, conventional semiparametric survival model; JM, joint model; LDH, lactate dehydrogenase; SLD, sum of the longest diameters of target lesions

coded in R, according to Equation (5), subsequently aggregated, and taken as an output from the *RsSimulx* script (see the Supplementary Materials, Section 3.1).

Similarly, these plots may be generated in R when using LME JM tools (see Figure S3). However, graphical diagnostics for multivariate JMs in *JMbayes* are limited. Thus, VPC plots for SLD and LDH as well as patient survival plots for the training data were generated for univariate models (JM SLD and JM LDH) handled within the *JM* package.

Finally, comparable goodness-of-fit (Figure 3 and S3) as well as parameter estimates (Table 2) were obtained for LME and NLME models against the training data. Upon achieving successful model convergence and adequate diagnostics, we considered further model applicability to predict biomarker dynamics as well as survival and perform discrimination analyses against the interim validation data. These steps are schematically described in Figure 1 (right column).

EXTERNAL VALIDATION

A validation procedure usually necessitates an analysis of model performance against an external data set—data that were not used in the earlier model qualification. As stated previously and represented in Figure 1, we performed such an analysis against an interim validation data set with biomarker and survival data known for the first 3 months of observation.

In *Monolix*, individual predictions of biomarker trends and patient survival were obtained via sampling from the conditional distribution of random effects, whereas the rest of the model parameters were fixed at their estimated values obtained during qualification. In LME JMs, these parameters were sampled from the posterior parameter distribution obtained from the qualification procedure.¹⁴

We further performed survival discrimination assessment for all generated LME and NLME JMs, thus comparing model predictions against actual survival outcomes in the validation study.

Survival discrimination performance is typically analyzed using the well-established metrics of area under the receiver operating characteristic curve (ROC-AUC) and BS (Brier score).^{53,54} ROC-AUC values range from 0 to 1, with a higher value representing higher discrimination performance, whereas a value of 0.5 would signify no discrimination ability. BS represents a mean square error of individual survival probability predictions, and its values range from 0 to 1, with lower values representing higher precision.⁵

Both ROC-AUC and BS were calculated for survival predictions $S_i(t)$ obtained at specific timepoints of interest (e.g., at different months following the longitudinal data cutoff in the validation study). Models with higher discrimination performance may potentially be further used to adequately stratify patient subgroups. Although this additional step is not described in the present tutorial, it has been detailed in a previously published JM analysis.⁵

The chosen data cutoff of 3 months in the validation data set represents an early biomarker assessment in selected NSCLC studies. It provides enough data to perform predictions of biomarker dynamics given the sample measurement frequency in the considered studies, combined with an appropriate JM structure (especially for NLME multivariate JMs—see Figure 5). However, depending on the clinical studies and data, a different cutoff time may be tested to identify the optimal amount of interim longitudinal data that would allow for adequate biomarker predictions and efficient survival discrimination.⁵

ROC-AUC and BS can be calculated in a unified manner for model predictions generated using LME and NLME JM tools, as shown in the Supplementary Materials, Section 4. Implemented ROC-AUC and BS calculations

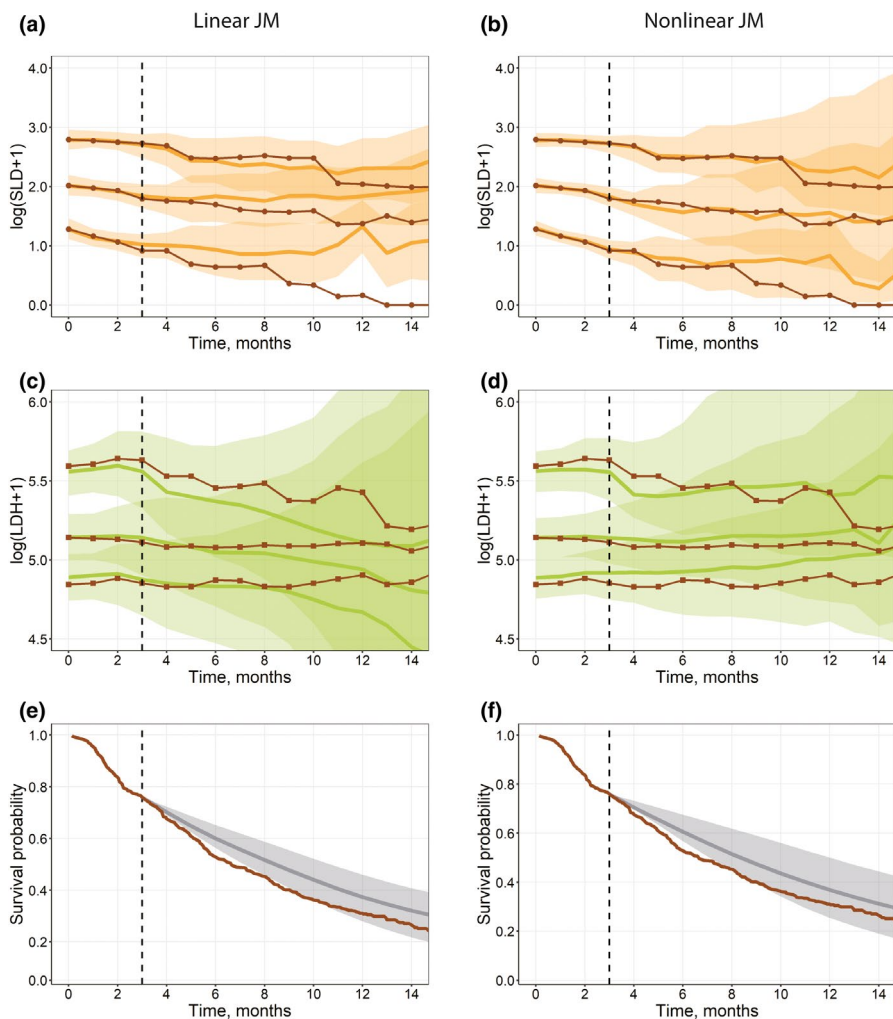


FIGURE 5 (a–d) Visual predictive check plots for the interim validation data set (Validation step), based on univariate linear joint models in *JM* (a, c) and a multivariate nonlinear biexponential joint model in *Monolix* (b, d). Solid black lines represent aggregated log-transformed biomarker data from the validation data set (circles, SLD; squares, LDH). Solid colored lines represent 10%, 50%, and 90% quantiles of averaged individual predictions (orange, SLD; green, LDH). Prediction intervals for each quantile are computed with a 95% confidence interval (shaded area). (e–f) Survival plot shows the experimental Kaplan-Meier curve (solid brown) from the validation data set and the mean model prediction (solid gray) with the interquartile range (shaded area) for a multivariate linear joint model in *JMbayes* (e) and a multivariate nonlinear biexponential joint model in *Monolix* (f). LDH, lactate dehydrogenase; SLD, sum of the longest diameters of target lesions

are adapted to right-censored data using the inverse probability of censoring weighting.^{55,56} Both ROC-AUC and BS were calculated using the means of sampled individual survival predictions without considering the uncertainty in model parameters and random effects. Also, the *JM* package includes embedded functions for ROC-AUC and BS calculations.¹⁴ Outputs of the discrimination analysis are presented in Figure 4.

The conventional semiparametric survival model (COX), which incorporated only baseline biomarker values, was a poor performer in these diagnostics, whereas the multivariate JMs in *JMbayes* and *Monolix* performed well, indicating that longitudinal trajectories of SLD and LDH biomarkers were important to achieve higher survival discrimination for the validation data. If further discrimination performance gain was sought, one may want to identify and add other baseline or longitudinal covariates deemed important to the augmented multivariate JMs. Interestingly, a higher number of significant baseline covariates in a conventional Cox proportional hazards model may result in a similar or higher discrimination performance versus JMs with a smaller number

of longitudinal and baseline biomarkers (data not shown). However, such baseline biomarker models would not allow for making inference for biomarker dynamics that may carry clinical importance or may even be foreseen as study endpoints. For all the univariate and multivariate survival models considered in this tutorial, an assessment should be performed to determine whether a biomarker's association parameter p value is below 0.05 (information usually available in the model summaries—refer to Table S3). Likewise, a relevant statistical criterion representing the goodness-of-fit and incorporating parameter count penalization (i.e., AIC) should be evaluated to compare the models and combat overparameterization.^{57,58}

To provide VPCs and patient survival prediction diagnostics (Figure 1, right column) using the selected LME and NLME JM tools, a set of functions and scripts such as those used in the model qualification steps were applied (see the Supplementary Materials, Sections 3 and 5). For LME JMs, which were implemented in the *JM* and *JMbayes* packages, the *predict()* and *survfitJM()* functions were used to obtain predictions of individual longitudinal trends and individual survival $S_i(t)$ for the validation data.

These functions represent outcomes of the sampling from subject-specific conditional probabilities of individual random effects and corresponding survival or biomarker predictions¹⁴ obtained from previously qualified JMs.

In Monolix, predictions of longitudinal biomarker trajectories and individual survival $S_i(t)$ were generated in the *lixoftConnectors/RsSimulx* packages (see the Supplementary Materials, Section 5.1). The estimated population parameters from previously qualified JMs were fixed and used for individual parameter sampling in a MCMC procedure⁵² against the interim validation data set.

For validation VPC plots, longitudinal biomarker predictions and $S_i(t)$ were simulated only for those patients who survived (event or censoring times for these patients should exceed the selected cutoff of 3 months), at time-points for which further experimental observations were available. The other features of the validation figures were the same as in the qualification step. VPCs and patient survival prediction plots were similarly generated for LME and NLME tools (Figure 5).

Within the tested JM tools, comparable diagnostic performance for OS predictions were achieved using the *JM/JMbayes* packages and Monolix. However, Monolix provided improved predictions of longitudinal biomarker trends. In the initial section of the present tutorial, we mentioned possible limitations of LME JMs, owing to the longer term behavior of the longitudinal submodels. Splines may cause nonmeaningful trends in the areas with few observations or extrapolation areas (in case of the validation data set, beyond the 3-month cutoff interval). Although these LME models provided an adequate goodness of fit versus NLME models in Monolix (e.g., compare Figure S3 and Figure 3) for the training data, the capability of LME JMs to adequately predict the longer term longitudinal biomarker dynamics for the validation data was limited (Figure 5).

We considered standard spline models in the *JM/JMbayes* packages and biexponential models in Monolix. Additional research may be required to determine whether these models are fully optimal compared with those that can be generated using customized knot positions for splines or using various empirical or mechanistic NLME biomarker submodels in Monolix. One option may also be to identify and introduce significant covariates for the longitudinal submodels of JMs (i.e., the biomarker trend may depend on patient's baseline characteristics, disease status, external conditions, chosen therapy, etc.). However, one needs to consider that these advanced JMs may require substantial computational times.

CONCLUSIONS

In this tutorial, we proposed and implemented a unified workflow for joint modeling to perform a comprehensive

analysis of multiple longitudinal biomarkers and a time-to-event endpoint. The workflow incorporates an initial data exploration phase, steps in performing JM building and qualification using a training data set, as well as testing against an external data set for model validation. It has been implemented in R-based LME tools and in more advanced NLME tools such as Monolix. These tools allow one to comprehensively assess information and data previously collected through clinical trials to predict the outcomes of another trial based on its interim data.

The outcomes of the analysis suggest that for the considered NSCLC data sets, nonlinear multivariate JMs in Monolix provide the highest performance in predicting longitudinal biomarker trends and survival discrimination. Moreover, advanced NLME joint modeling tools such as NONMEM, Monolix, or Stan may further overcome actual study data limitations with respect to the clinical development process. When multiple biomarkers are being evaluated in patients it often happens that both the measurement schedule and the duration of observations vary across biomarkers, hence the biomarker data are not aligned in time. For the sake of simplicity, in this tutorial, we considered a scenario whereby all interim biomarker data were collected up to the same cutoff in time. However, the approach we delineated may also be applied to an analysis that includes multiple biomarkers that have different observation times.

JMs are emerging as a data analytics methodology, which may provide a link between advanced mechanistic biomarker submodels (which may include quantitative systems pharmacology submodels) and time-to-event data to perform state-of-the-art clinical trial simulations. Joint modeling, in effect, represents a powerful methodology in pharmacometric analyses, supported by efficient tools, to characterize the dynamic behavior of biomarkers and its association with clinical trial outcomes.

ACKNOWLEDGMENTS

The clinical study data used in this tutorial were derived from www.projectdatasphere.org, which is maintained by Project Data Sphere LLC. Neither Project Data Sphere LLC nor the owners/sponsors of the original data have contributed to or are in any way responsible for the content of the analyses presented in this article.

CONFLICT OF INTEREST

Kirill Zhudnikov, Sergey Gavrillov, Alina Sofronova, Oleg Stepanov, Nataliya Kudryashova, and Kirill Peskov are employees of M&S Decisions, LLC. Kirill Peskov is also affiliated with the Sechenov First Moscow State Medical University. Gabriel Helmlinger is an employee of Obsidian Therapeutics.

REFERENCES

- Kurtz DM, Esfahani MS, Scherer F, et al. Dynamic risk profiling using serial tumor biomarkers for personalized outcome prediction. *Cell*. 2019;178(3):699-713.
- US Department of Health and Human Services, Food and Drug Administration. Multiple endpoints in clinical trials. Guidance for industry. <https://www.fda.gov/files/drugs/published/Multi-Endpoints-in-Clinical-Trials-Guidance-for-Industry.pdf>. Published 2017. Accessed April 15, 2021.
- Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol*. 2016;16(1):117. <https://doi.org/10.1186/s12874-016-0212-5>
- Zhudenkov K, Palmér R, Jauhiainen A, et al. Longitudinal FEV1 and exacerbation risk in COPD: quantifying the association using joint modelling. *Int J Chron Obstruct Pulmon Dis*. 2021;16:101-111. <https://doi.org/10.2147/COPD.S284720>
- Gavrilov S, Zhudenkov K, Helmlinger G, Dunyak J, Peskov K, Aksenov S. Longitudinal tumor size and neutrophil-to-lymphocyte ratio are prognostic biomarkers for overall survival in patients with advanced non-small cell lung cancer treated with durvalumab. *CPT Pharmacometrics Syst Pharmacol*. 2021;10(1):67-74. <https://doi.org/10.1002/psp4.12578>
- Desmée S, Mentré F, Veyrat-Follet C, Sébastien B, Guedj J. Using the SAEM algorithm for mechanistic joint models characterizing the relationship between nonlinear PSA kinetics and survival in prostate cancer patients: joint model for nonlinear kinetics and survival data. *Biometrics*. 2017;73(1):305-312. <https://doi.org/10.1111/biom.12537>
- Chesnaye NC, Tripepi G, Dekker FW, Zoccali C, Zwinderman AH, Jager KJ. An introduction to joint models—applications in nephrology. *Clin Kidney J*. 2020;13(2):143-149. <https://doi.org/10.1093/ckj/sfaa024>
- Mchunu NN, Mwambi HG, Reddy T, Yende-Zuma N, Naidoo K. Joint modelling of longitudinal and time-to-event data: an illustration using CD4 count and mortality in a cohort of patients initiated on antiretroviral therapy. *BMC Infect Dis*. 2020;20(1):256. <https://doi.org/10.1186/s12879-020-04962-3>
- Wu Y, Zhang X, He Y, et al. Predicting Alzheimer's disease based on survival data and longitudinally measured performance on cognitive and functional scales. *Psychiatry Res*. 2020;291:113201. <https://doi.org/10.1016/j.psychres.2020.113201>
- Tardivon C, Desmée S, Kerioui M, et al. Association between tumor size kinetics and survival in patients with urothelial carcinoma treated with atezolizumab: implication for patient follow-up. *Clin Pharmacol Ther*. 2019;106(4):810-820. <https://doi.org/10.1002/cpt.1450>
- Muse R, Romero K, Stafford B, Larkindale JA. *Multivariate Joint Model for Predicting End-Stage Renal Disease in Polycystic Kidney Disease (ESRD)*. Critical Path Institute; 2020.
- Morales JF, Lingineni K, Thomas M, et al. *Modeling Longitudinal glucose dynamics to inform type 1 diabetes prevention clinical trial studies*. American Conference on Pharmacometrics; 2020.
- Harhay MO, Gasparini A, Walkey AJ, et al. Assessing the course of organ dysfunction using joint longitudinal and time-to-event modeling in the vasopressin and septic shock trial. *Crit Care Explor*. 2020;2(4):e0104. <https://doi.org/10.1097/CCE.000000000000104>
- Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data*. Chapman and Hall/CRC; 2012.
- Natale RB, Thongprasert S, Greco FA, et al. Phase III trial of vandetanib compared with erlotinib in patients with previously treated advanced non-small-cell lung cancer. *JCO*. 2011;29(8):1059-1066. <https://doi.org/10.1200/JCO.2010.28.5981>
- Herbst RS, Ansari R, Bustin F, et al. Efficacy of bevacizumab plus erlotinib versus erlotinib alone in advanced non-small-cell lung cancer after failure of standard first-line chemotherapy (BeTa): a double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2011;377(9780):1846-1854. [https://doi.org/10.1016/S0140-6736\(11\)60545-X](https://doi.org/10.1016/S0140-6736(11)60545-X)
- Benzekry S, Lamont C, Beheshti A, et al. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput Biol*. 2014;10(8):e1003800. <https://doi.org/10.1371/journal.pcbi.1003800>
- Schober P, Vetter TR. Survival analysis and interpretation of time-to-event data: the tortoise and the Hare. *Anest Analg*. 2018;127(3):792-798. <https://doi.org/10.1213/ANE.00000000000003653>
- Yin A, Moes DJAR, Hasselt JGC, Swen JJ, Guchelaar H. A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(10):720-737. <https://doi.org/10.1002/psp4.12450>
- Nguyen THT, Mouksassi M, Holford N, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst Pharmacol*. 2017;6(2):87-109. <https://doi.org/10.1002/psp4.12161>
- Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An overview of joint modeling of time-to-event and longitudinal outcomes. *Ann Rev Stat Appl*. 2019;6(1):223-240. <https://doi.org/10.1146/annurev-statistics-030718-105048>
- Rizopoulos D. *Joint models for longitudinal and survival data*. Erasmus Summer Program; 2019. <http://www.drizopoulos.com/courses/EMC/ESP72.pdf>. Accessed October 15, 2020.
- Alam K. Joint modeling of longitudinal continuous, longitudinal ordinal, and time-to-event outcomes. *Lifetime Data Anal*. 2021;27:64-90.
- CRAN Documentation. JM package: joint modeling of longitudinal and survival data. <https://cran.r-project.org/web/packages/JM/index.html>. Published 2020. Accessed April 15, 2021.
- CRAN Documentation. JMbayes package: joint modeling of longitudinal and time-to-event data under a Bayesian approach. <https://cran.r-project.org/web/packages/JMbayes/index.html>. Published 2020. Accessed April 15, 2021.
- CRAN Documentation. rstanarm package: Bayesian applied regression modeling via Stan. <https://cran.r-project.org/web/packages/rstanarm/index.html>. Published 2020. Accessed May 14, 2021.
- Pillai G, Mentré F, Steimer J-L. Non-linear mixed effects modeling – from methodology and software development to driving implementation in drug development science. *J Pharmacokinetic Pharmacodyn*. 2005;32(2):161-183. <https://doi.org/10.1007/s10928-005-0062-y>
- Bauer RJ. NONMEM Tutorial Part I: description of commands and options, with simple examples of population analysis. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(8):525-537. <https://doi.org/10.1002/psp4.12404>

29. Monolix 2020R1 Documentation. <https://monolix.lixoft.com>. Published 2020. Accessed April 15, 2021.
30. Traynard P, Ayrat G, Twarogowska M, Chauvin J. Efficient pharmacokinetic modeling workflow with the MonolixSuite: a case study of remifentanyl. *CPT Pharmacometrics Syst Pharmacol*. 2020;9(4):198-210. <https://doi.org/10.1002/psp4.12500>
31. Brilleman S, Crowther M, Moreno-Betancur M, Novik JB, Wolfe R. *Joint longitudinal and time-to-event models via Stan*. Paper presented at: Pacific Grove, CA, USA: Stancon Conference 2018. https://www.sambrilleman.com/pdf/conference/2018_StanCon_notebook.pdf. Accessed January 10-12, 2018.
32. Riglet F, Mentre F, Veyrat-Follet C, Bertrand J. Bayesian individual dynamic predictions with uncertainty of longitudinal biomarkers and risks of survival events in a joint modelling framework: a comparison between Stan, Monolix, and NONMEM. *AAPS J*. 2020;22(2):50. <https://doi.org/10.1208/s12248-019-0388-9>
33. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020. <https://www.R-project.org/>. Accessed May 15, 2021.
34. Project Data Sphere. Project data sphere repository. <https://www.projectdatasphere.org/>. Published 2020. Accessed April 15, 2021.
35. Zhang X, Guo M, Fan J, et al. Prognostic significance of serum LDH in small cell lung cancer: a systematic review with meta-analysis. *Cancer Biomark*. 2016;16(3):415-423. <https://doi.org/10.3233/CBM-160580>
36. Forkasiewicz A, Dorociak M, Stach K, Szelachowski P, Tabola R, Augoff K. The usefulness of lactate dehydrogenase measurements in current oncological practice. *Cell Mol Biol Lett*. 2020;25(1):35. <https://doi.org/10.1186/s11658-020-00228-7>
37. Schwartz LH, Litière S, de Vries E, et al. RECIST 1.1—Update and clarification: from the RECIST committee. *Eur J Cancer*. 2016;62:132-137. <https://doi.org/10.1016/j.ejca.2016.03.081>
38. Monolix 2020R1 Online Documentation. Defining a dataset in Monolix. <https://monolix.lixoft.com/data-and-models/creating-data-set>. Published 2021. Accessed April 15, 2021.
39. Lixoft. Lixoft software. <https://lixoft.com/>. Accessed April 15, 2021.
40. Monolix 2020R1 Online Documentation. Time-to-event modeling with the MonolixSuite, part 1: introduction. <https://monolix.lixoft.com/case-studies/time-event-modeling-monolixsuite-part-1-introduction/>. Published 2021. Accessed April 15, 2021.
41. Monolix 2020R1 Online Documentation. R functions to run Monolix. <https://monolix.lixoft.com/monolix-api/>. Published 2021. Accessed April 15, 2021.
42. CRAN Documentation. nlme package: linear and nonlinear mixed effects models. CRAN.R. <https://cran.r-project.org/web/packages/nlme/index.html>. Published 2021. Accessed April 15, 2021.
43. CRAN Documentation. lme4 package: mixed-effects models in R. CRAN.R. <https://www.r-project.org/nosvn/pandoc/lme4.html>. Published 2020. Accessed April 15, 2021.
44. CRAN Documentation. Survival package: survival analysis. CRAN.R. <https://cran.r-project.org/web/packages/survival/index.html>. Published 2021. Accessed April 15, 2021.
45. Duffull SB, Wright DFB, Winter HR. Interpreting population pharmacokinetic-pharmacodynamic analyses—a clinical viewpoint: Interpretation of population analyses for clinicians. *Br J Clin Pharmacol*. 2011;71(6):807-814. <https://doi.org/10.1111/j.1365-2125.2010.03891.x>
46. Singh R, Mukhopadhyay K. Survival analysis in clinical trials: basics and must know areas. *Perspect Clin Res*. 2011;2(4):145. <https://doi.org/10.4103/2229-3485.86872>
47. Monolix 2020R1 Online Documentation. Time-to-event data models. <https://monolix.lixoft.com/data-and-models/ttedata/>. Published 2021. Accessed April 15, 2021.
48. Monolix 2020R1 Online Documentation. Understanding shrinkage and how to circumvent it. <https://monolix.lixoft.com/faq/understanding-shrinkage-circumvent/>. Published 2021. Accessed April 15, 2021.
49. Monolix 2019 user guide. <https://monolix.lixoft.com/single-page/>. Published 2020. Accessed April 15, 2021.
50. Samson A, Lavielle M, Mentré F. Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Comput Stat Data Anal*. 2006;51(3):1562-1574. <https://doi.org/10.1016/j.csda.2006.05.007>
51. Owen JS, Fiedler-Kelly J. *Introduction to Population Pharmacokinetic/Pharmacodynamic Analysis with Nonlinear Mixed Effects Models*. Wiley; 2014.
52. Monolix 2020R1 Online Documentation. Conditional distribution. <https://monolix.lixoft.com/tasks/conditional-distribution/>. Published 2020. Accessed April 15, 2021.
53. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
54. Brier GW. Verification of forecasts expressed in terms of probability. *US Weather Bureau, Washington, D C*. 1950;78:1-3.
55. Gerds TA, Schumacher M. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biom J*. 2006;48(6):1029-1040. <https://doi.org/10.1002/bimj.200610301>
56. Desmée S, Mentré F, Veyrat-Follet C, Sébastien B, Guedj J. Nonlinear joint models for individual dynamic prediction of risk of death using Hamiltonian Monte Carlo: application to metastatic prostate cancer. *BMC Med Res Methodol*. 2017;17(1). <https://doi.org/10.1186/s12874-017-0382-9>
57. Liang H, Zou G. Improved AIC selection strategy for survival analysis. *Comput Stat Data Anal*. 2008;52(5):2538-2548. <https://doi.org/10.1016/j.csda.2007.09.003>
58. Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*. 1987;52(3):345-370. <https://doi.org/10.1007/BF02294361>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Zhudenkov K, Gavrilov S, Sofronova A, et al. A workflow for the joint modeling of longitudinal and event data in the development of therapeutics: Tools, statistical methods, and diagnostics. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:425-437. doi:[10.1002/psp4.12763](https://doi.org/10.1002/psp4.12763)