

Guideline for data and codes

Using the SAEM algorithm for mechanistic joint models characterizing the relationship between nonlinear PSA kinetics and survival in prostate cancer patients

Solène Desmée, France Mentré, Christine Veyrat-Follet,
Bernard Sébastien and Jérémie Guedj

Contents

1	Data	1
2	MLXTRAN code and Monolix use	2
2.1	Main code	2
2.2	Structural model	3
2.3	Running Monolix	3
3	R code	4

1 Data

The training dataset is named JM400.txt. Column names are:

- TIME: time from treatment initiation (negative before treatment initiation)
- ID: patient's identifiants
- $Y = \log(\text{PSA}+1)$ if YTYPE equals to 1
= vital status (1 if death, 0 if censored) if YTYPE equals to 2
- CENS = 1 if PSA is below the limit of quantification (0.1 ng.ml^{-1}) and 0 otherwise
- YTYPE: type of observation, 1 for PSA, 2 for vital status
- TIME_ENDtr: covariate that indicates the day of end of treatment

2 MLXTRAN code and Monolix use

2.1 Main code

We used Monolix version 4.3.2 on Matlab R2013a. We provide a folder `Project_JM.mlxtran` which specifies the project features and the different paths. It is structured as followed:

- **DESCRIPTION:** comments on the project.
- **DATA:** information from the dataset `JM400.txt`. The column names of the present dataset are already keywords for Monolix. The path needs to be set.
- **INDIVIDUAL:** specifies the probability distribution of the individual parameters and the inter-individual variability.
- **STRUCTURAL_MODEL:** indicates the external file which contains the structural model, here `Structural_model_finalJM.txt` (see details in the paragraph 2.2). You have to change the path.
- **OBSERVATIONS:** specifies the statistical model of the observation model. `y1` corresponds to the longitudinal $\log(\text{PSA}+1)$ observations and the error model is constant. `survival` corresponds to the time-to-event.
- **TASKS:** specifies the algorithms settings and the tasks to perform.
 - All algorithms settings are specified in the available `Project_JM_algorithms.xmlx` file. The main settings to examine here are:
 - * The seed can be changed (`<algorithms seed="123456">` line 2).
 - * The number of iterations in the E- and M steps of SAEM can be fixed (`<iop_Kauto value="0,0"/>` line 25 and `<vna value="500,200"/>` line 22) or automatic (`<iop_Kauto value="1,1"/>` and `<vna value="500,200"/>` indicates that the maximum number of iterations is 500 for the E-step and 200 for the M-step). We recommend to successively run several population parameters estimation in order to be sure to not have a local extremum. Moreover SAEM convergence can be visually checked throughout the iterations.
 - * We choose 1 Markov chain (`<nmc value="1"/>` line 28) after trying 3 chains without change in results.
 - * To obtain precise estimation of the log-likelihood (small s.e.), we increased the Monte-Carlo size for LL to 200,000 (`<M_is value="200000"/>` line 4).
 - Results folder can be chosen. Here `%MLXPROJECT%` indicates that result folder is the same than the `mlxtran` file.
 - The tasks are:
 - * Estimation of the population parameters: Initializations of fixed effects (`pop_`), standard deviation of the random effects (`omega_`) and residual error (`a_`) parameters have to be specified. SAEM is not very sensitive to initial values. But for illustration and rapidity, we proposed here values closed to the final values.
 - * Estimation of the Fisher Information matrix (and standard errors): in case of joint modelling, Fisher Information Matrix has to be computed by stochastic approximation.

- * Estimation of the individual parameters: we use the conditional modes (i.e., maximization of the conditional probabilities $p(\Phi_i|y_i;\hat{\theta})$). The file "indiv_parameters.txt" is created at this moment and will be useful to plot the figures presented in the article and whose R codes are described in the section 3.
- * Estimation of the log-likelihood: in case of joint modelling, likelihood has to be estimated by important sampling.

2.2 Structural model

The `Structural_model_finalJM.txt` is structured by blocks as followed:

- **DESCRIPTION:** comments on the model
- **INPUT:** vector of parameters to estimate
- **EQUATION:** description of the mathematical model
 t is the keyword for the time regressor.
`odeType=stiff` indicates that Monolix have to use the solver for stiff ODE system. Monolix will use the CVODE solver for stiff ODEs (Cohen and Hindmarsh, 1996).
Here we can fix parameters, specify constraints and transformations.
`PSA_0`, `S_0` and `R_0` are the initial values of the ODE system, i.e., the values of PSA, S and R respectively at first time of PSA observation.
`ddt_S`, `ddt_R` and `ddt_PSA` are the derivative of S, R and PSA respectively with respect to time t .
- **OBSERVATION:** The probability distribution of the event data is defined with the hazard function.
`type=event` specify that the data is of type "event".
- **OUTPUT:** vector of the output of the model

2.3 Running Monolix

The project can be directly run using the following commands in Matlab:

```
monolix;
aMonolixProject = MonolixProject('Path/to/Project_JM.mlxtran');
aMonolixProject.load('Path/to/Project_JM.mlxtran');
aMonolixProject.run()
```

Once the project has been created, one can directly open the project from the Monolix interface in order to change the model, the data or the algorithm calibration in a user-friendly manner.

3 R code

After the Monolix code has been run the R code can be used to generate the figures of the manuscript. The following needs to be done beforehand:

- install the `survival` and `deSolve` packages
- change the working directory (i.e., where Monolix outputs were saved) in `setwd()` (line 18) and the directory to save figures in `way` (line 21)
- change the values of the parameters to the value found in the file "estimated.txt" generated by Monolix (lines 25-27 and 29)
- import the individual parameters provided by Monolix (line 36)
- import the data (line 43)

RCodes.R provides the R codes to reproduce the Figures:

2. Individual fits of PSA kinetics and hazard function (line 116). Other samples of patients can be chosen (line 121).
3. Residuals for the longitudinal part (IWRES) and the survival part (Cox-Snell and Martingale residuals) (line 234)
4. Estimated Kaplan-Meier survival curve and its confidence interval and mean survival curve (line 453)