

# Guideline for simulated data and codes

## A joint analysis proposal of nonlinear longitudinal and time-to-event right-, interval-censored data for modeling pregnancy miscarriage

Rolando De la Cruz, Marc Lavielle, Cristian Meza and Vicente Núñez-Antón

We include as supplementary material a simulated dataset and the associated files to replicate this study with Monolix:

- `jointProject-data.csv`: the simulated data.
- `jointProject-simulation.mlxtran`: the MLxtran model.
- `jointProject-simulation.mlxproperties`: properties of the plots. It allows the project to reload all the graphical properties to be able to reproduce exactly the same plots (in terms of color, split by categories,...).
- `model-simulation.txt`: the structural model defined in Equation 3 of Section 3 of the paper.

Below we provide details and guidelines to run this project in Monolix.

### 1 Data

Since the data are confidential, we simulated longitudinal profiles disturbing the real dataset by adding a noise. In Figure 1, we display the simulated dataset.

The simulated data file is named `jointProject-data.csv` with the following column names:

- `id`: patient's identifiers.
- `time`: observation time in days.
- `y`:  $\log_{10}(\beta - HCG)$  if `ytype` equals to 1, or, TTE data if `ytype` equals to 2.
- `group`: covariate group equals to 0 if woman belongs to normal group or 1 if she belongs to abnormal group.
- `ytype`: type of observation, 1 for longitudinal data, 2 for TTE data.
- `mdv`: identifier to ignore the observation `y` information of that line (use here to select the abnormal group in the TTE part).

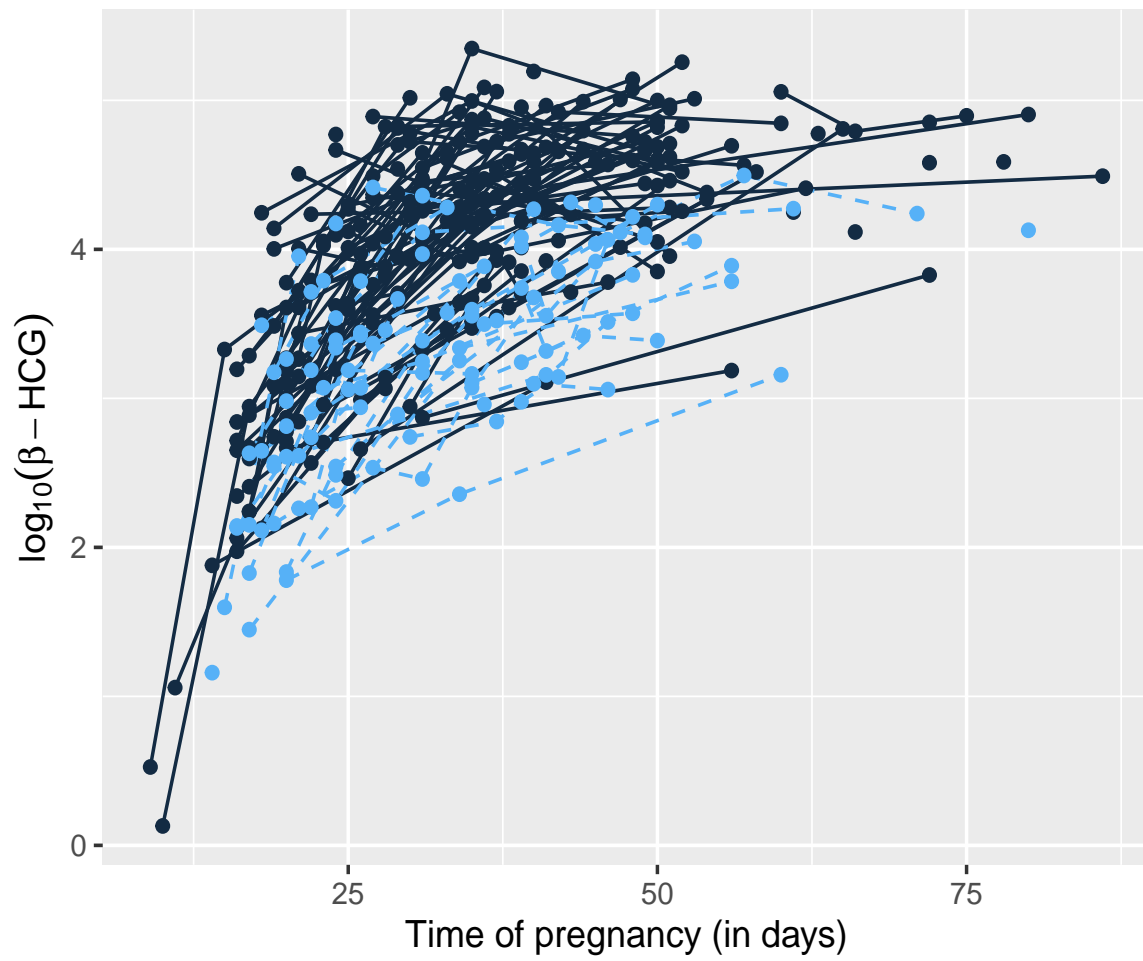


Figure 1: Simulated time profiles for normal (in darkblue lines) and abnormal pregnancies (in lightblue dashed lines).

## 2 Monolix and Mlxtran models

We use the Monolix software version 2021R1 to fit this simulated dataset. The project associated to the Model  $\mathcal{M}_2$  is resumed in the following Mlxtran project called

`jointProject-simulation.mlxtran`

with the following main structures:

- `<DATAFILE>`: data information such as file name of the dataset, delimiter, header titles, and type of the columns.
- `<MODEL>`: definition of the `[COVARIATE]`, `[INDIVIDUAL]` and `[LONGITUDINAL]` sections, this last indicating the structural model used.
- `<FIT>`: indicate the data information.
- `<PARAMETER>`: initial values used in the SAEM algorithm.
- `<MONOLIX>`: definition of the main tasks such as the estimation of the population parameters, estimation of the Fisher Information matrix and standard errors, estimation of the individual parameters or estimation of the log-likelihood among others.

In Monolix the covariate and statistical models for the individual parameters is provided through the user interface.

The Structural model defined in `model-simulation.txt` is structured following these blocks:

- `[LONGITUDINAL]`: In this section, the structural model and the observation model (error model) are described.
  - `input = { }`: a list which defines the parameters that are estimated.
- `EQUATION`: equations are implemented in this block defining the parametric models for longitudinal data and TTE part.
- `DEFINITION`: this block defines the TTE part indicating the interval censoring.
- `OUTPUT`: this block declares variables whose values are exported for the various tasks.

To execute this project, open `jointProject-simulation.mlxtran` in Monolix and run it.