Mixed-effects parameter estimation using Monolix

Introduction

Nonlinear mixed-effects modeling is widely used in the field of pharmacometrics (the quantitative analysis of pharmacological data), both in academia and in the industry. The general purpose is to capture *population* characteristics of dynamical processes such as pharmacokinetics and pharmacodynamics of a drug. These population parameters can then be used to improve *individual* predictions. Pharmacokinetics (PK) is usually referred as to what the body does to a drug and comprises the processes of absorption, distribution, metabolism and elimination. Pharmacodynamics deals with what the drug does to the body, i.e. the toxicity and efficacy. One can think of two global clinical objectives of population PK/PD modeling studies: 1) to design dosing and scheduling in clinical trials (to ensure safety and efficacy for a large majority of the population a description of inter-individual variability is needed) and 2) to individualize dosing regimen through bayesian estimation (that combines population and individual information obtained by therapeutic drug monitoring). For the latter, identification of individual *covariates* (such as the sex, the weight or the age) that impact on the model parameters is of particular relevance.

In this session, we will see how to use the Monolix software to develop a mixed-effects model for description of the pharmacokinetics of warfarin, an anti-coagulant drug used in the prevention of thrombosis and thromboembolism (the formation of blood clots in the blood vessels and their migration elsewhere in the body, respectively). A single large loading dose of warfarin sodium, 1.5 mg/kg of body weight, was administered orally to all 32 subjects. The data comes from a study by [O'Reilly and Aggeler, 1968].

1. Load the data and launch Monolix

1) Create a folder for the hands-on session

2) Download the data at the following link:

`http://benzekry.perso.math.cnrs.fr/DONNEES/tp_monolix/warfarin_data_pk.txt`

3) Open Monolix by typing the following command in a terminal:

`> monolix.sh`
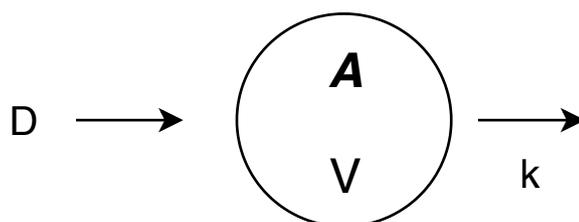
Use the following activation key: 7484-4558-4904-9809.

4) Create a new project

5) Load the data. The column "AMT" corresponds to the dose of drug that was administered, the column "CONC" to the plasmatic concentrations that were measured. Look at the "spaghetti plot" (DATAVIEWER).

6) Use a log scale to have a first guess at how many elimination phases there are in the data?

2. No absorption

Assume the simplest PK model for elimination of the drug, that is linear first-order kinetics.

This model can be written in terms of a differential equation as:

$$\begin{cases} \frac{dA}{dt} = -kA \\ A(t = 0) = D \end{cases} \quad C(t) = \frac{A(t)}{V},$$

where $C(t)$ is the concentration (in mg·L^{-1}), V is the volume of distribution of the drug (in L), used to convert the administered *dose* (in mg) into a *concentration*, and k is the elimination

rate (in hour^{-1}). Note that pharmacologists prefer to use the more meaningful notion of *clearance*, defined by $Cl = kV$ and thus in units of L·hour^{-1}, which expresses what volume of blood is "cleared" from the drug by unit of time.
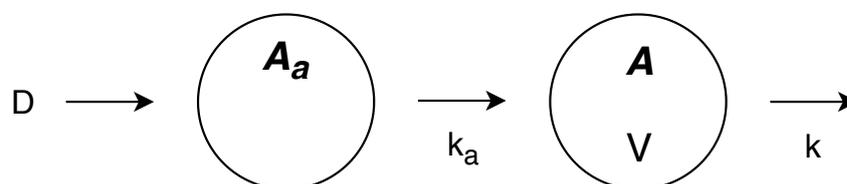
0) Download the structural model at:

`http://benzekry.perso.math.cnrs.fr/DONNEES/tp_monolix/one_comp_bolus.txt`

1) Go to structural model and load the file `one_comp_bolus.txt` as model.

2) Explore initial values of the parameters to get a reasonable initial guess

3) Launch the fit. (Do you understand the meaning of all possible tasks? (EBE = Empirical Bayes Estimates))

4) Look at the goodness-of-fit of the model for individual fits (plots and Observations vs. Predictions). What do you conclude about this model?

5) Compute likelihood. Write down AIC (Akaike Information Criterion) for future comparison Save this project.

3. Absorption

In fact, as seen directly on the data, the dose does not get directly to the plasma. This is due to the mode of administration of the drug, which was by oral take. Thus, the drug has to go through an absorption compartment (the gastro-intestinal system) before reaching the systemic circulation.

The equations for this model write, with $A_a(t)$ and A the amounts of drug in the absorption and central compartments, respectively, and $C(t)$ the concentration in the central (systemic)

compartment:

$$\begin{cases} \frac{dA_a}{dt} = -k_a A_a \\ \frac{dA}{dt} = k_a A_a - kA \\ A_a(t=0) = D, \quad A(t=0) = 0 \end{cases} \qquad C(t) = \frac{A(t)}{V}.$$

1) Write this model in MLXTRAN in a .txt file. MLXTRAN is the language used by Monolix. You can use the previous model file as a template.

Note the use of the `depot` function of MLXTRAN to deal with administration of the doses. The role of this function is to inject doses in the variables, which would normally require Dirac functions (if no administration length (i.e. perfusion time)) in the differential equations. It does nothing more than $A(t_D^+) = A(t_D^-) + D(t_D)$, where t_D is a dose administration time and $D(t_D)$ is the dose given at this time (defined in the column AMT (for amount) in the data).

2) Use a proportional error model. Look at the formulas for the error and individual models. Compare with the theoretical formulas from the class.

3) *Fit*

    3.1) Launch estimation of the population and individual parameters. Does the model seem appropriate on individual fits?

    3.2) What individual contributes most to the likelihood? Run likelihood estimation task, then plot individual contributions (parameters in the "Plots" button)? Does this correspond with individual fits?

Note: each time you run a task and save in Monolix, files are added to a folder with the name of the project containing the results. You can find a summary of the fit, population and individual parameters, predictions for each of the individuals, etc... All these information is stored in .txt files and can easily be loaded from external softwares (such as R, python or matlab) for further analysis.

4) *Residuals*

    4.1) Plot the individual and population residuals (parameters in the "Plots" button). Plot only conditional mode ("Display" panel on the right). (**Do you understand what is the

difference between the conditional distribution plotted and the conditional mode?)

4.2) Add the population residuals (see "Subplots" tab on the right).

4.3) Do you see misspecification? Are they symmetrically distributed around zero? Do they seem gaussian? Where does this come from (in the fits)?

4.4) Look at the statistical tests on the residuals ("Results" tab). Look also at the "Distribution of residuals" plots. Do they confirm what you were observing?

4.5) What parameter gives the (proportional) error between the model fit and the data? What is its value?

5) *Population goodness-of-fit*

5.1) Plot the Visual Predictive Check (VPC) and Prediction Distribution ("Plots" button in "Statistical model and tasks"). Add the data and predicted percentiles to the VPC and prediction distribution. Interpret. Does the model describe well the population variability?

5.2) What are the colored blue and orange areas in the VPC?

5.3) Is AIC improved compared to the first model without absorption?

6) *Population parameters. Practical identifiability*

6.1) What parameter has largest inter-individual variability (standard deviation of the random effects)? What parameter corresponds to the latter?

6.2) Launch estimation of standard errors ("Statistical model and Tasks"). (**Do you understand what does "Use linearization method" means and what it has to do with standard errors?) What parameter has largest uncertainty in its estimation? Of how much? What could improve identifiability of this parameter? (Note: you can also plot standard errors as an option in the "Plots" task.)

6.3) What is the difference between standard errors and standard deviations of the random effects?

6.4) Is there any noticeable correlation between the estimates?

7) *Individual parameters*

(** 7.0) How are the individual parameters estimated in Monolix?)

7.1) Look at the distribution of the estimated individual parameters ("Plots" tab). Do the estimated individual parameters seem to follow their theoretical (estimated) distribution? How is the latter defined?

(** 7.1 bis) What is η-shrinkage? Get the values by ticking "Information" in subpanel. Are they good or bad?)

7.2) Do the distributions of random effects seem symmetrical and gaussian?

7.3) Is there any noticeable correlation between the individual parameters (i.e. random effects)? Why is this different from the question 5.4? ("Plots>Correlation between random effects tab, General>Information subtab")

(7.4) What happens if you put a full correlation (tick all the parameters in the "Tasks" tab)? Is AIC improved? What happens on the standard errors? Why?)

4. Covariates

1) *Data exploration*

1.1) Go back to the data viewer. Stratify by the sex covariate. Assign colors. Do you see difference in the kinetics between sexes?

1.2) Dichotomize the weight variable. Is there difference by weight?

1.3) Plot weight versus sex. What do you see?

1.4) By plotting sex against sex, determine how many females are present in the data.

2) *Model for the covariates*

2.1) Look at the "Individual parameters vs covariates" plot. Which covariates seem to impact on what parameters?

2.2) Confirm by looking at the statistical tests performed between covariates and random effects ("Results" tab).

2.3) Add this covariate to the model. Look at how it impacts on the formula defining the individual model. Save the project with new name first. Use the last estimates as initial values for accelerating convergence ("Initial estimates" tab).

2.4) Look at the value of the β parameter in the population estimates. Is it significant? (** What statistical test was used to test the null hypothesis $\beta = 0$?)

2.5) What was the effect of adding the covariate on the (unexplained) inter-individual variability? Interpret.

2.6) Look at the individual parameters vs covariates. Tick random effects (why?) in the "Display" subtab. Is the effect of the covariate still present?

2.7) How is the VPC? Has it improved? Has AIC improved?

5*. Global identifiability

To test whether the parameter estimated by likelihood maximization were global maxima, Monolix has a built-in function to perform the estimation several times from random initial conditions. This is the "Assessment" button, below "Run scenario" in "Tasks". Try this functionality, possibly by choosing yourself the bounds of the parameter space to explore. Interpret the results in terms of global identifiability.

Note: Don't hesitate to look at Monolix's documentation and case studies (`http://monolix.lixoft.com/`) to go further into population modeling. In particular, there are examples on how to estimate individual parameters using a bayesian approach (of specific relevance when combined to mixed-effects modeling). Monolix can also be used for other data types (such as time-to-event). There are also R equivalents of the tasks performed by Monolix (such as likelihood maximization for population estimation, see packages `mlxR`, which is linked to Monolix and only accepts MLXTRAN-written models and `saemix`, which does not require any Monolix installation). Monolix projects can also be combined to the R package `simulx` for simulations in a population context (such as clinical trial design).

References

[O'Reilly and Aggeler, 1968] O'Reilly, R. A. and Aggeler, P. M. (1968). Studies on coumarin anticoagulant drugs. Initiation of warfarin therapy without a loading dose. *Circulation*, 38(1):169–177.