

Projet Pharmacométrie

Vivien DUPONT – Maxime BEAULIEU

L'ensemble des analyses et de la programmation est disponible au lien suivant : [Maitre-Vivi1/Pharmaco \(github.com\)](https://github.com/Maitre-Vivi1/Pharmaco)

I) Modélisation de la dynamique tumorale

Question 1

Mesures

49 patients ont participé à l'étude. Les patients ont en moyenne effectué 2.7 scanners. Le temps médian de suivi est de 67 jours. 15 patients n'ont effectué qu'un seul scanner. Deux patients n'ont pas effectué un scanner initial (en $t=0$). Les patients ont en moyenne reçu 6 doses de médicament entre le premier scanner et le dernier.

Âge

En moyenne, les patients ont entre 45 et 83 ans (66 ans en moyenne et 9.5 ans d'écart-type).

Hémoglobine

En moyenne, les patients ont une hémoglobine de 121 g/l (de 88 au minimum jusqu'à 156 et 17.8 en écart-type).

Les valeurs d'hémoglobines sont plutôt faibles car si l'on regarde les recommandations de l'OMS¹ nous avons :

- Femmes qui ne sont pas enceintes (à partir de 15 ans) : 120 g/l ou plus
- Hommes (à partir de 15 ans) : 130 g/l ou plus

Métastases

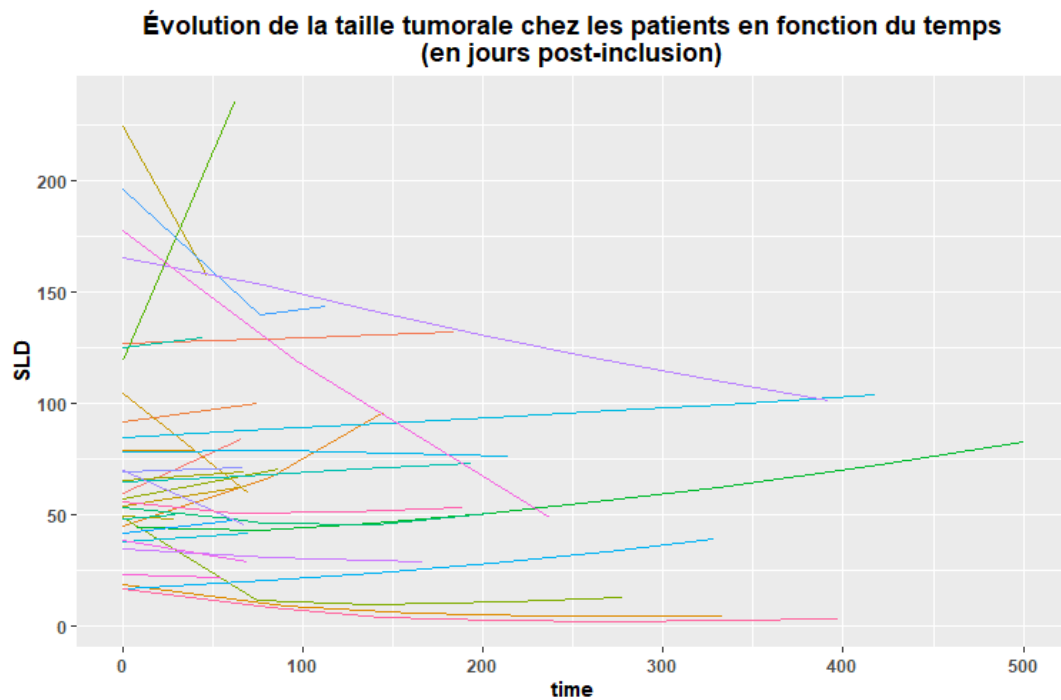
11 patients sur 49 souffrent de métastases au foie.

Sexe

Le sexe des patients est absent de la base de données. Ce choix est dommageable pour la qualité de la modélisation.

¹ Recommandations niveau d'hémoglobine OMS :
https://www.who.int/vmnis/indicators/haemoglobin_fr.pdf

Représentation graphique



La somme des diamètres des tumeurs semble dans un premier temps décroître chez beaucoup de patients puis recroître passé un certain seuil. Les dynamiques de croissance et de décroissance ne semblent pas linéaires. Un minimum de SLD est atteint entre 50 et 200 jours chez la plupart des patients.

Question 2

Aucunes données dans la littérature -- à notre connaissance -- ne permettent d'estimer un des paramètres du modèle. Dans le cas contraire, on aurait pu se servir de ces valeurs.

Estimation de BSLD

BSLD est le niveau de SLD au premier jour d'inclusion. BSLD peut alors être estimé comme la moyenne de SLD au temps 0. L'écart-type de BSLD peut également être estimé de cette manière.

Ainsi :

$$\text{BSLD} = 74$$

$$\text{BSLD_sd} = 50$$

Modèle de Wang

Le modèle de Wang décrit une dynamique de décroissance exponentielle et une dynamique de croissance linéaire en fonction du temps.

Afin d'estimer g et d , il faut séparer la base de données en deux : une partie contenant les observations où SLD croît et une autre partie où SLD décroît.

Ainsi l'estimation de g peut se réaliser en réalisant le modèle linéaire mixte suivant dans la base de données où SLD est en croissance :

$$SLD_{ij} = \beta_0 + \gamma_{0i} + (\beta_1 + \gamma_{1i}) \times t_j$$

Où SLD_{ij} est l'observation de SLD de l'individu i au temps j.

g correspond alors à l'estimation de β_1 . L'écart-type de g correspond à l'écart-type de β_1 .

Par conséquent :

$$g_Wang = 0.09$$

$$g_sd_Wang = 0.05$$

L'estimation de d se réalise grâce à la base de données dans laquelle SLD décroît. On transforme la variable SLD en log afin de recouvrir le coefficient d. On ajuste le modèle linéaire mixte suivant :

$$\log(SLD_{ij}) = \beta_0 + \gamma_{0i} + (\beta_1 + \gamma_{1i}) \times t_j$$

d correspond à $|\beta_1|$. L'écart-type de d est estimé de la même manière.

Par conséquent :

$$d_Wang = 0.004$$

$$d_sd_Wang = 0.0009$$

Modèle de Stein-Fojo

d et g sont estimés de manière similaire que dans le modèle de Wang à l'exception près que g est contenu dans une exponentielle et le modèle linéaire mixte est adapté en conséquence. De plus, le -1 dans la formule ne va influencer que sur le niveau initial de SLD, il n'a donc aucun impact sur l'estimation de g et d.

Par conséquent, d aura la même estimation que celui dans le modèle de Wang. Pour g, il suffit d'appliquer le même modèle linéaire mixte que pour d mais en se plaçant sur la base de données où SLD croît.

On a donc :

$$g_SteinFojo = 0.002$$

$$g_sd_SteinFojo = 0.0006$$

et

$$d_SteinFojo = 0.004$$

$$d_sd_SteinFojo = 0.0009$$

Modèle de Chatterjee

Le modèle de Chatterjee fait intervenir un paramètre de convexité ϕ . Celui-ci représente le poids de la décroissance par rapport à la croissance. Plus ϕ est élevé et plus la décroissance est importante par rapport à la croissance. Afin de mesurer l'importance de la décroissance par rapport à la croissance,

nous avons mesuré les écart-type de SLD chez les différents sujets dans les bases de données de croissance puis de décroissance puis nous en avons fait la somme.

Dans la base de données de croissance la moyenne d'écart-type de SLD chez les patients est de 15 tandis que dans la base de données de décroissance, ce dernier est de 24. On peut estimer ϕ comme suit :

$$\phi = \frac{24}{24 + 15}$$
$$\phi = 0.62$$

Les estimations de d et g sont réalisées de la même manière que pour le modèle de Stein-Fojo puisqu'il représente les mêmes tendances de décroissance et de croissance.

On a donc :

$$g_Chatterjee = 0.002$$

$$g_sd_Chatterjee = 0.0006$$

et

$$d_Chatterjee = 0.004$$

$$d_sd_Chatterjee = 0.0009$$

Remarque

Pour chacun des modèles, les estimations des paramètres sont assez proches des valeurs initiales fournies dans l'algorithme.

Question 3

La croissance tumorale chez les patients étant à tendance non-linéaire le modèle de Wang ne semble pas bien ajuster les données au vu du graphique de la question 1. Le modèle structural de Wang semble alors le moins pertinent. Le modèle de Chatterjee permet de prendre en compte une importance différente de la croissance de la tumeur par rapport à sa décroissance au niveau populationnel et individuel. Sans lancer les modèles, le modèle structural le plus pertinent est celui de Chatterjee.

Néanmoins, on compare les estimations et la qualité des trois modèles. Les paramètres initiaux utilisés sont ceux estimés à la question 2. Ces paramètres ont été transformés log-normalement afin de garder des valeurs positives. C'est également le cas de la variable réponse SLD qui a été transformée log-normalement pour rester positive.

Les modèles d'erreurs des modèles ont d'avoir été supposés constant puis, le cas échéant transformés si c'était nécessaire. Pour chacun des modèles, un graphique observations vs prédictions a été généré. Ce graphique a permis d'apprécier si les erreurs étaient proportionnelles aux prédictions. Si c'était le cas, le modèle d'erreur était ajusté. Au final, tous les modèles d'erreurs ont été paramétré comme constants.

Les résultats des modèles sont très semblables. Notamment pour les paramètres d et g qui étaient très ressemblant dans la plupart des modèles. Cette observation valide notre stratégie d'estimation des paramètres initiaux.

Les différents modèles sont comparés sans tenir compte des éventuelles interactions entre effets aléatoires ou de ses derniers avec des covariables. Les modèles sont tout de même comparables puisque les paramètres estimés ont tous un rapport valeur-écart-type inférieur à 50%.

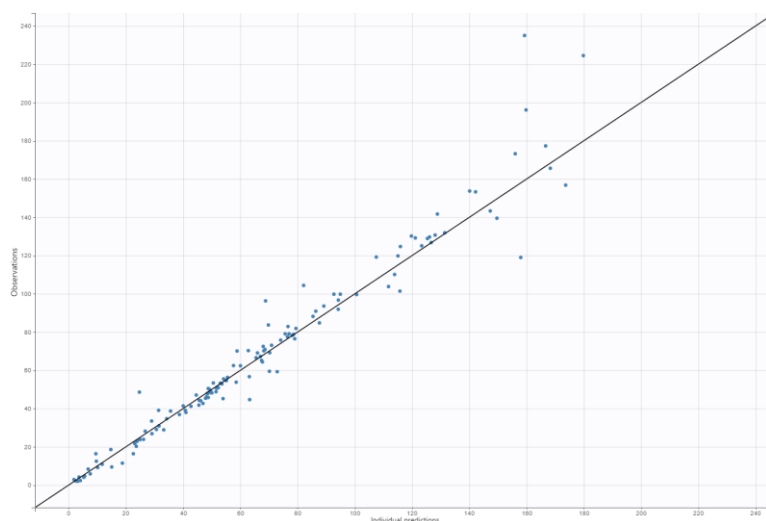
Tableau 1: Comparaison des paramètres

Paramètres de population	Transformation	Wang	Stein-Fojo	Chatterjee
BSLD	Lognormale	60.42	60.90	60.03
d	Lognormale	0.0041	0.004	0.004
g	Lognormale	0.0600	0.0011	0.0017
ϕ	Lognormale	-	-	0.41

Afin de sélectionner le meilleur modèle, on compare les BICc des trois modèles. Le BICc le plus faible est obtenu par le modèle de Chatterjee (BICc = 1213).

Ce dernier est bien ajusté aux données :

Figure 1: Ajustement du modèle de Chatterjee



Question 4

On choisit de procéder à une procédure de sélection des interactions pas à pas. On commence avec le modèle de base de Chatterjee estimé à la question 3 puis on ajoute l'interaction la plus significative d'une covariable (s'il y en a une de significative). Puis, on ajoute encore une fois l'interaction la plus significative dans le nouveau modèle et on vérifie que la première interaction n'est pas devenue non-significative (le cas échéant on la supprimerait). On s'arrête jusqu'à n'avoir plus aucune variable à ajouter ou retirer.

La procédure sélectionne une seule covariable dans le modèle : La présence de métastase dans le foie à la baseline sur le niveau de SLD à la baseline.

Les nouveaux paramètres de populations sont :

Paramètres	BSLD	d	g	ϕ	BSLD_BLIVER
Estimations	54.26	0.004	0.0017	0.4	0.45

Le modèle complet est disponible en annexe : Tableau 2 & Figure 2.

Question 5

Le temps à laquelle la valeur de SLD est minimale dans le meilleur modèle est obtenu en annulant la dérivée du modèle par rapport au temps. Le modèle atteint bien un unique minimum puisque strictement convexe (somme de fonctions strictement convexes).

$$SLD'(t) = BSLD \times (-d \times \phi \times e^{-d \times t} + (1 - \phi) \times g \times e^{g \times t})$$

$$SLD'(t) = 0 \Leftrightarrow$$

$$BSLD \times (-d \times \phi \times e^{-d \times t} + (1 - \phi) \times g \times e^{g \times t}) = 0 \Leftrightarrow$$

$$(1 - \phi) \times g \times e^{g \times t} = \phi \times d \times e^{-d \times t} \Leftrightarrow$$

$$e^{(g+d) \times t} = \frac{d \times \phi}{(1 - \phi) \times g} \Leftrightarrow$$

$$t^* = \frac{\ln(d \times \phi) - \ln((1 - \phi) \times g)}{g + d}$$

Ainsi, $SLD(t)$ est minimisée en t^* .

Par les données précédentes, $t^* = 79$. Le traitement permet une minimisation de la somme des diamètres des tumeurs au 79^{ème} jour depuis le début du traitement.

II) Modélisation de la cinétique virale du SARS-CoV-2

Question 1

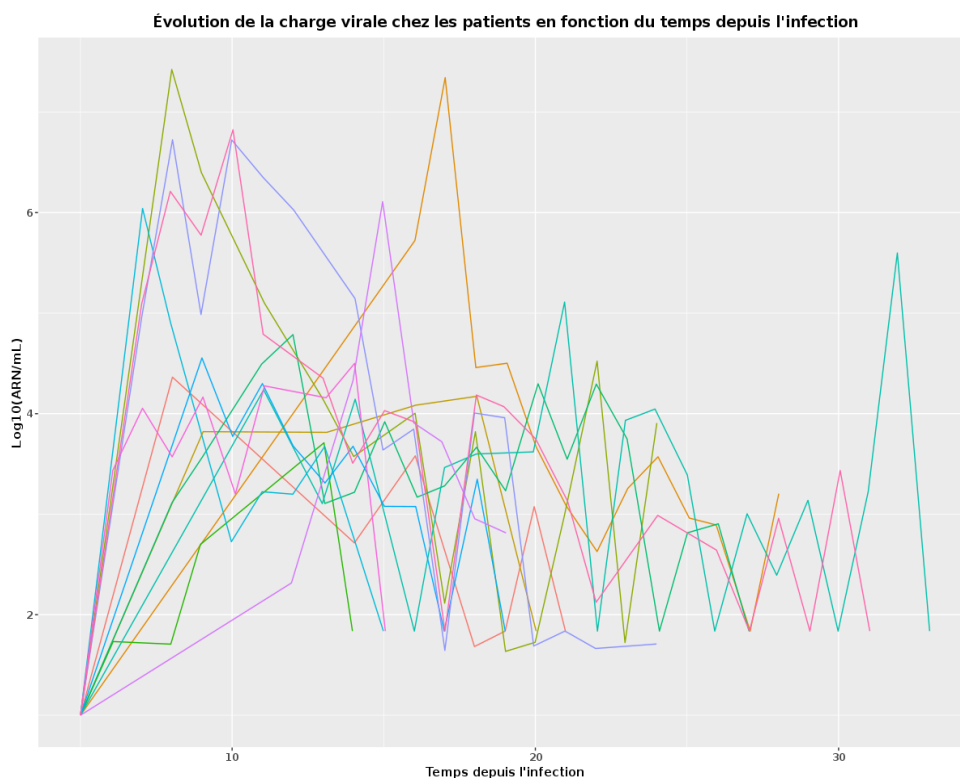
Notre base de données est constituée de quatre variables, l'identifiant de chaque participant à l'étude, une variable donnant le temps depuis les premiers symptômes du COVID-19, une variable représentant le temps depuis l'infection (5 jours avant les premiers symptômes) et une dernière variable, la charge virale du patient (à l'échelle log10).

```
> summary(data_vl$id) # 13 patients
 2  3  4  6  7  8  9 11 12 14 16 17 18
 7 13  5 14  5 19 21  7 11 16  7  9 22
> 156/13 # 12 mesures par patient en moyenne (min = 7 ; max = 22)
[1] 12
```

Nous pouvons voir que nous avons 13 identifiants dans la base de données, avec en moyenne 12 mesures par patient et un minimum et maximum, respectivement de 7 et 22 mesures.

```
> summary(data_vl$LVL)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.635   2.723   3.569   3.538   4.167   7.424
```

Nous remarquons que lors de cette étude, les charges virales observées chez les patients étaient comprises entre le log10 de 1,635 ARN/mL (=0,2135 ARN/mL) et le log10 de 7,424 ARN/mL (=0,8706 ARN/mL).



Ici nous pouvons voir la charge virale (à l'échelle log10) des 13 patients suivis. Malgré une variabilité inter-sujet non-négligeable, nous pouvons déceler une tendance croissante tout de suite après l'infection, et cela pendant près d'une semaine, puis décroissante.

Question 2

a)

- β est le taux d'infection, c'est une sorte de coefficient (unité : cellules infectées par une cellule par ml déjà malade).
- δ le taux d'élimination du virus (unité : proportion/pourcentage).
- p est le nombre de nouveaux virus produit par une cellule infectée, durant toute sa durée de vie (unité : ARN/mL).

b) La sélection du modèle se fait par le biais de l'analyse de plusieurs tableaux et graphiques donnés par Monolix. Tout d'abord, nous pouvons regarder le tableau des paramètres estimés en regardant le pourcentage de variabilité des estimateurs, en cherchant un pourcentage le plus faible possible. Ensuite, nous pouvons regarder la matrice variance-covariance des paramètres à estimer pour savoir s'il y a des interactions entre les paramètres estimés. De plus, le graphique des Observations vs Prédictions peut être une bonne indication visuelle si le modèle prédit correctement les données (les points doivent suivre le plus fidèlement possible la droite). Finalement, le critère BIC est un indicateur très important pour juger la qualité d'un modèle et pour comparer des modèles emboîtés, on préférera, au vu de cet indicateur, le modèle avec le BIC le plus faible.

Nous avons défini les valeurs initiales des paramètres par tâtonnement, en regardant quelles valeurs (plausibles) permettaient de correspondre au mieux au nuage de points de chaque individu. Nous nous sommes principalement intéressés à l'individu 18, qui comptabilise le plus de mesures de 0 à 30 jours après l'infection. Nous avons donc conclu à ces valeurs initiales :

- $\beta = 5$
- $\delta = 0,5$
- $p = 0,9$

Cf : Figure 3

Le modèle que nous avons sélectionné est un modèle avec une corrélation entre chacun des trois paramètres à estimer (après avoir consulté la matrice variance-covariance), avec une erreur proportionnelle, des effets aléatoires pour tous les paramètres à estimer, une distribution log-normale pour ces trois paramètres. Les estimateurs obtenus sont les suivants :

- $\hat{\beta} = 31,95$ avec une erreur standard égale à 10,82 (34%)
- $\hat{\delta} = 0,18$ avec une erreur standard égale à 0,036 (20%)
- $\hat{p} = 0,02$ avec une erreur standard égale à 0,016 (8%)

Cf : Figure 4

Ces erreurs standards sont convenables.

On ne rejette pas l'hypothèse de normalité des effets aléatoires à partir du test de Shapiro-Wilk (p-valeur très largement non-significative). De même, nous ne rejetons pas l'hypothèse de normalité des résidus. Finalement, nous pouvons dire que le modèle sélectionné est améliorable, il ne s'adapte pas vraiment aux observations.

Cf : Figure 5 & Tableau 3 & Tableau 4

Question 3

Le traitement antiviral administré dès l'arrivée des premiers symptômes permet de réduire p , la production de virus par les cellules infectées, tout en conservant le taux d'élimination du virus au même niveau. Cela n'aura donc pour effet qu'une diminution plus rapide du nombre de virions dans le corps de chaque patient, à partir des premiers symptômes. Si le pic de charge virale a déjà eu lieu (c'est à dire qu'il arrive moins de cinq jours après l'infection), alors ce traitement n'a aucun effet sur ce dernier, sinon cela réduirait la charge virale maximale mesurée chez le patient. De plus, comme la charge virale est croissante pendant les premiers jours suivants l'infection, puis décroissante, le traitement n'a qu'un effet négatif sur la quantité de virions dans le corps et donc cela réduirait le temps nécessaire pour la charge virale soit indétectable.

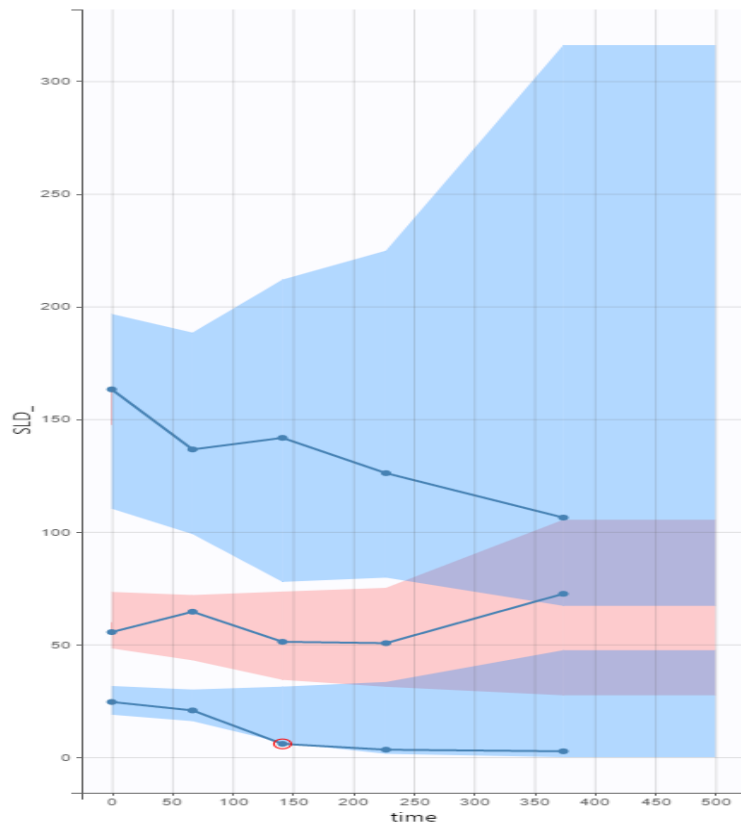
NB : Le modèle permettant de modéliser ce traitement est disponible dans le dépôt git « viral_kinetic_antiviral ».

Annexe
Exercice 1

Tableau 2: Modèle final

	VALUE	STOCH. APPROX.	
		S.E.	R.S.E.(%)
Fixed Effects			
BSLD_pop	54.26	10.78	19.9
beta_BSLD_BLIVER_TRUE	0.45	0.23	51.4
d_pop	0.004	0.000013	0.322
g_pop	0.0017	0.000013	0.798
phi_pop	0.4	1.05	265
Standard Deviation of the Random Effects			
omega_BSLD	0.64	0.079	12.4
omega_d	0.0013	0.00084	67.1
omega_g	0.0024	0.0066	281
omega_phi	0.64	2.15	334
Error Model Parameters			
a	0.22	0.039	17.9

Figure 2 : visual predictive check du modèle final



Exercise 2

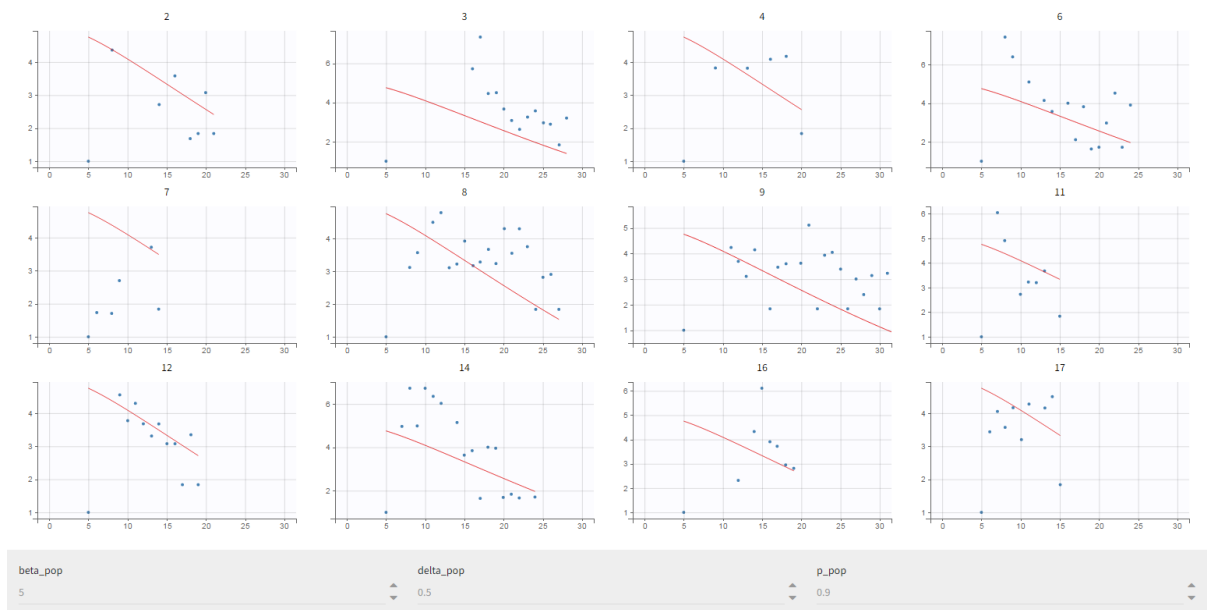


Figure 3: Fit individuals

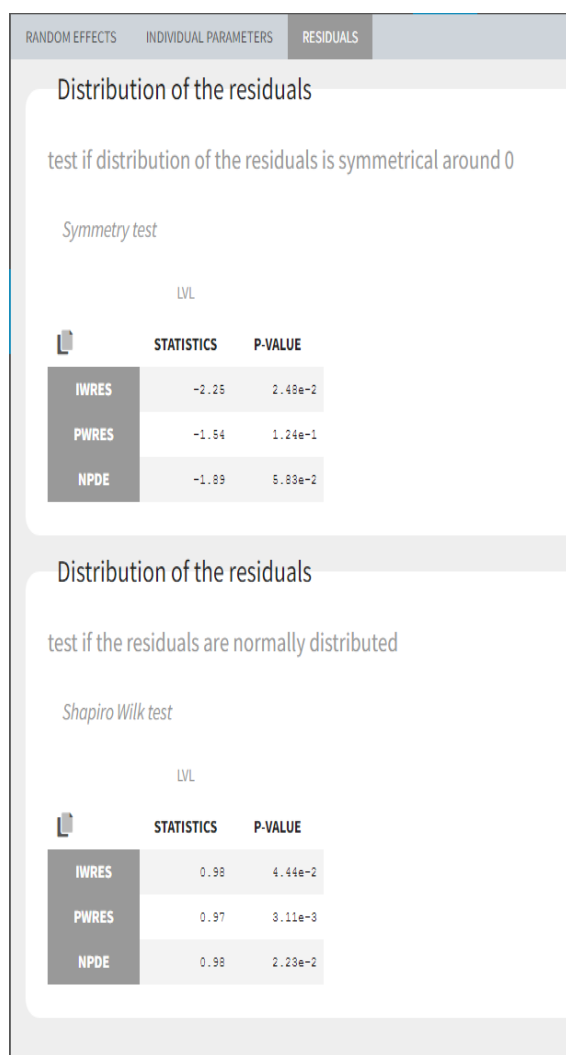


Figure 4: Tests sur la normalité des résidus et celle des effets aléatoires

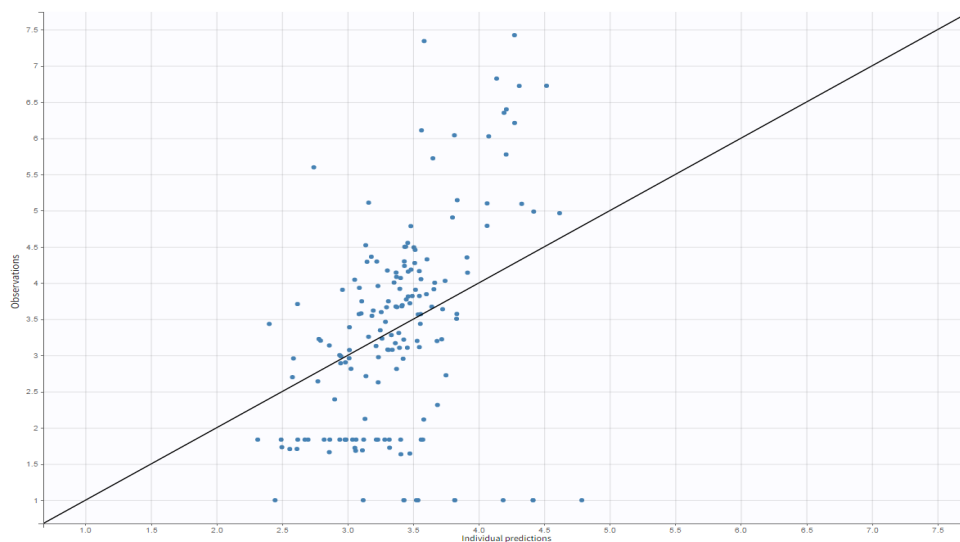


Figure 5: Observations vs Prédictions

Population parameter estimates

	VALUE	STOCH. APPROX.	
		S.E.	R.S.E.(%)
Fixed Effects			
beta_pop	31.95	10.82	33.9
delta_pop	0.18	0.036	20.4
p_pop	0.02	0.016	79.2
Standard Deviation of the Random Effects			
omega_beta	0.92	0.027	2.91
omega_delta	0.54		
omega_p	2.22		
Correlations			
corr_delta_beta	-1	0.0038	0.380
corr_p_beta	-1		
corr_p_delta	1	0.0033	0.335
Error Model Parameters			
b	0.38	0.023	6.08

Elapsed time (seconds): 197

Tableau 3 : Paramètres estimés

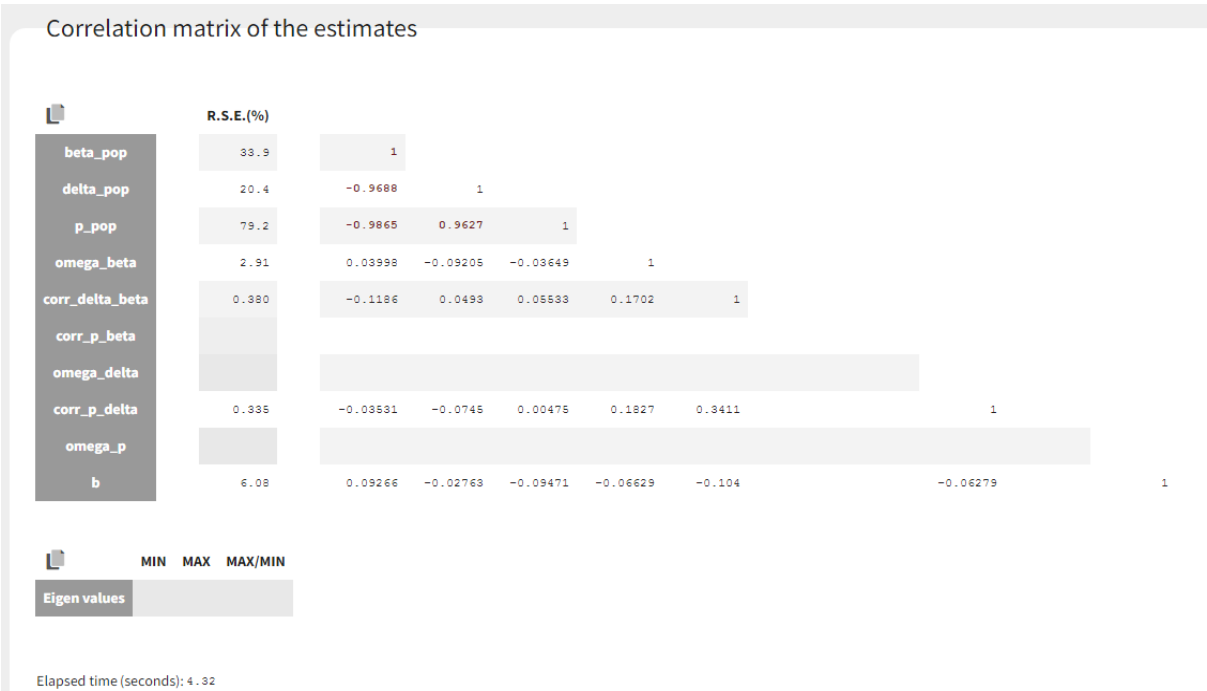


Tableau 4: Matrice des corrélations