MIXTURE MODELS AND MODEL MIXTURES WITH MONOLIX

Marc Lavielle [(1)], Hector Mesa [(1)],
Kaelig Chatel [(1)], An Vermeulen [(2)]

(1) INRIA Saclay

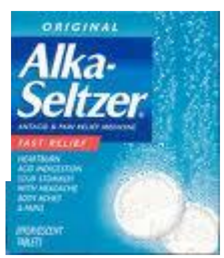
(2) J & J Pharmaceutical R & D

Some mixtures…

Yesterday night…

Some mixtures…

or this morning…



Outline

- **Mixture distributions**
  - Supervised learning
  - Unsupervised learning
- **Model mixtures**
  - Between subject model mixture
  - Within subject model mixture
- **The methodology**
- **Conclusions**

Mixture models

N subjects, $i = 1,2,\ldots,N$
K groups, $k = 1,2,\ldots,K$

y_i vector of observations from subject i

ψ_i vector of individual parameter of subject i

z_i label (categorical covariate): $z_i \in \{1,2,\ldots,K\}$

$$\text{If } z_i = k, \text{ then } (y_i,\psi_i) \sim p_k(y_i,\psi_i)$$

Mixture models

$$\boxed{p_k(y_i,\psi_i) = p_k(\psi_i)\,p_k(y_i|\psi_i)}$$

1) Mixture of distributions

$$\boxed{\text{If } z_i = k, \text{ then } \psi_i \sim p_k(\psi_i)}$$

2) Mixture of models

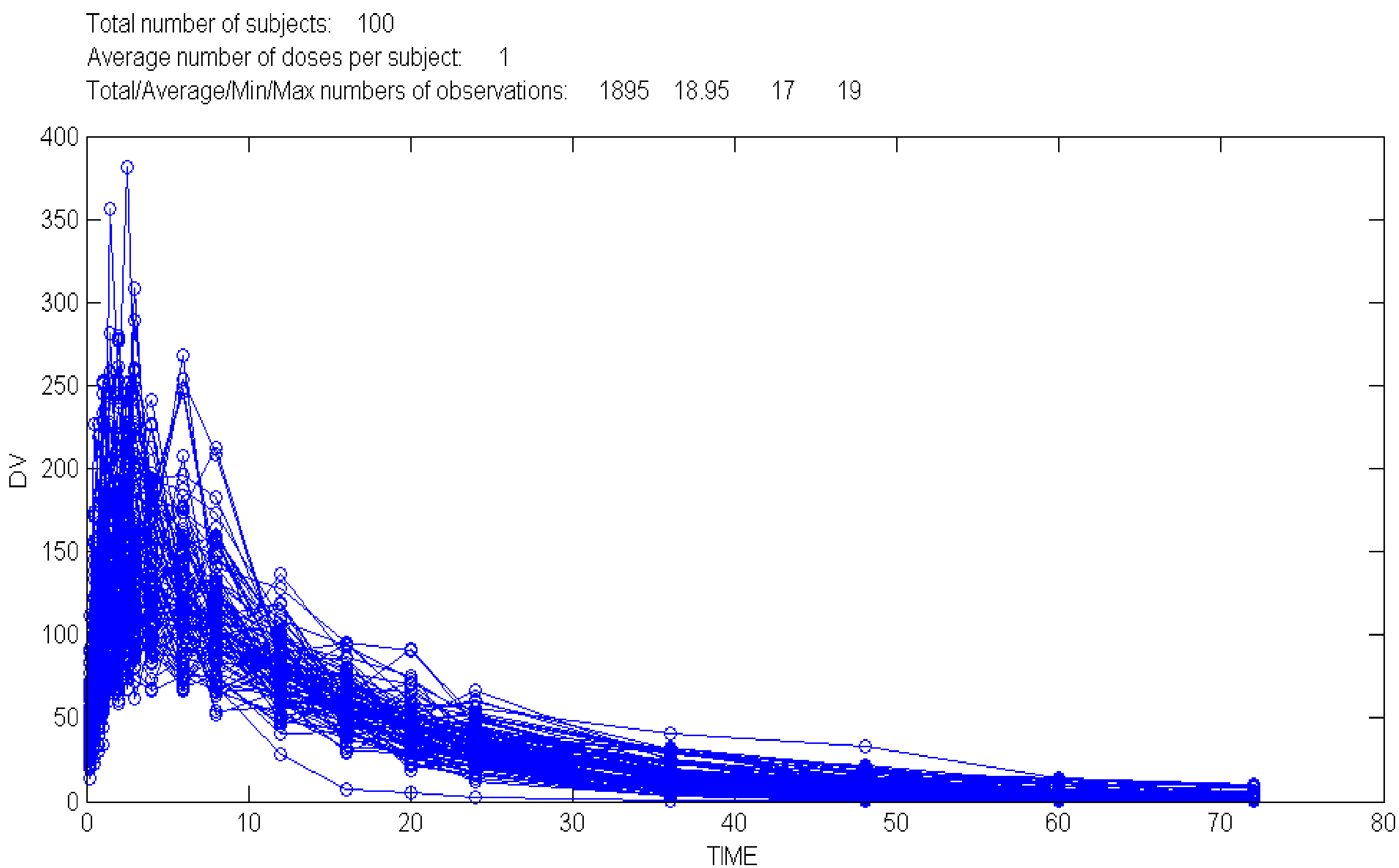
$$\boxed{\text{If } z_i = k, \text{ then } y_i|\psi_i \sim p_k(y_i|\psi_i)}$$

I

Mixture of distributions

Mixture of distributions

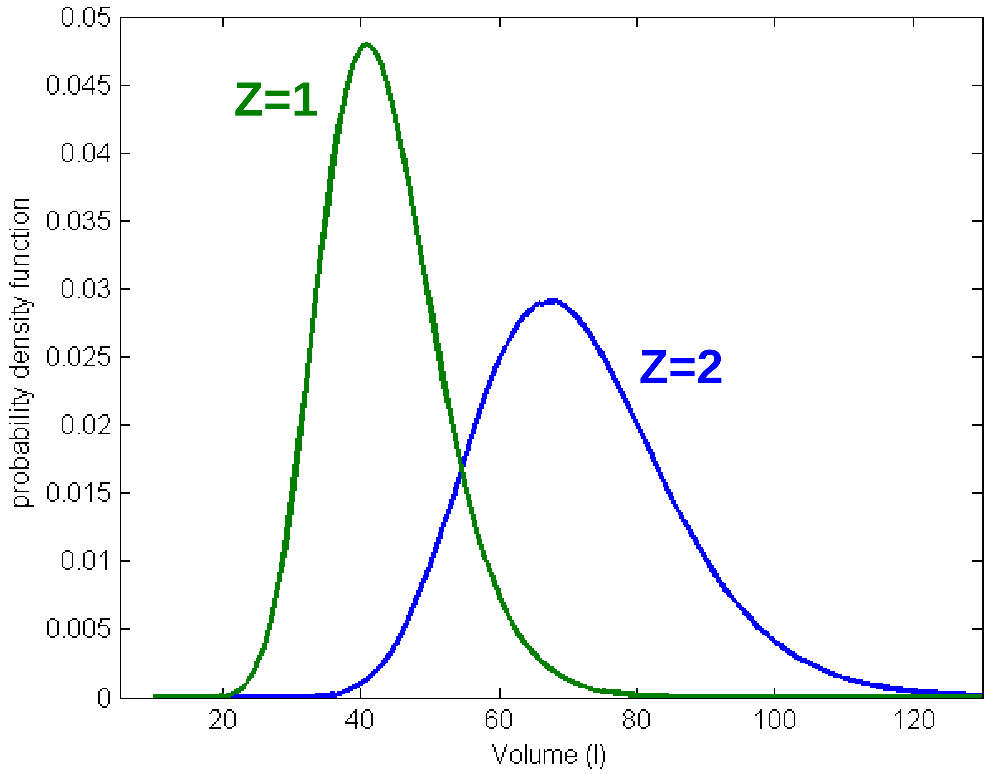
Exemple: $\psi_i = (ka_i, Vol_i, Cl_i)$



Mixture of distributions

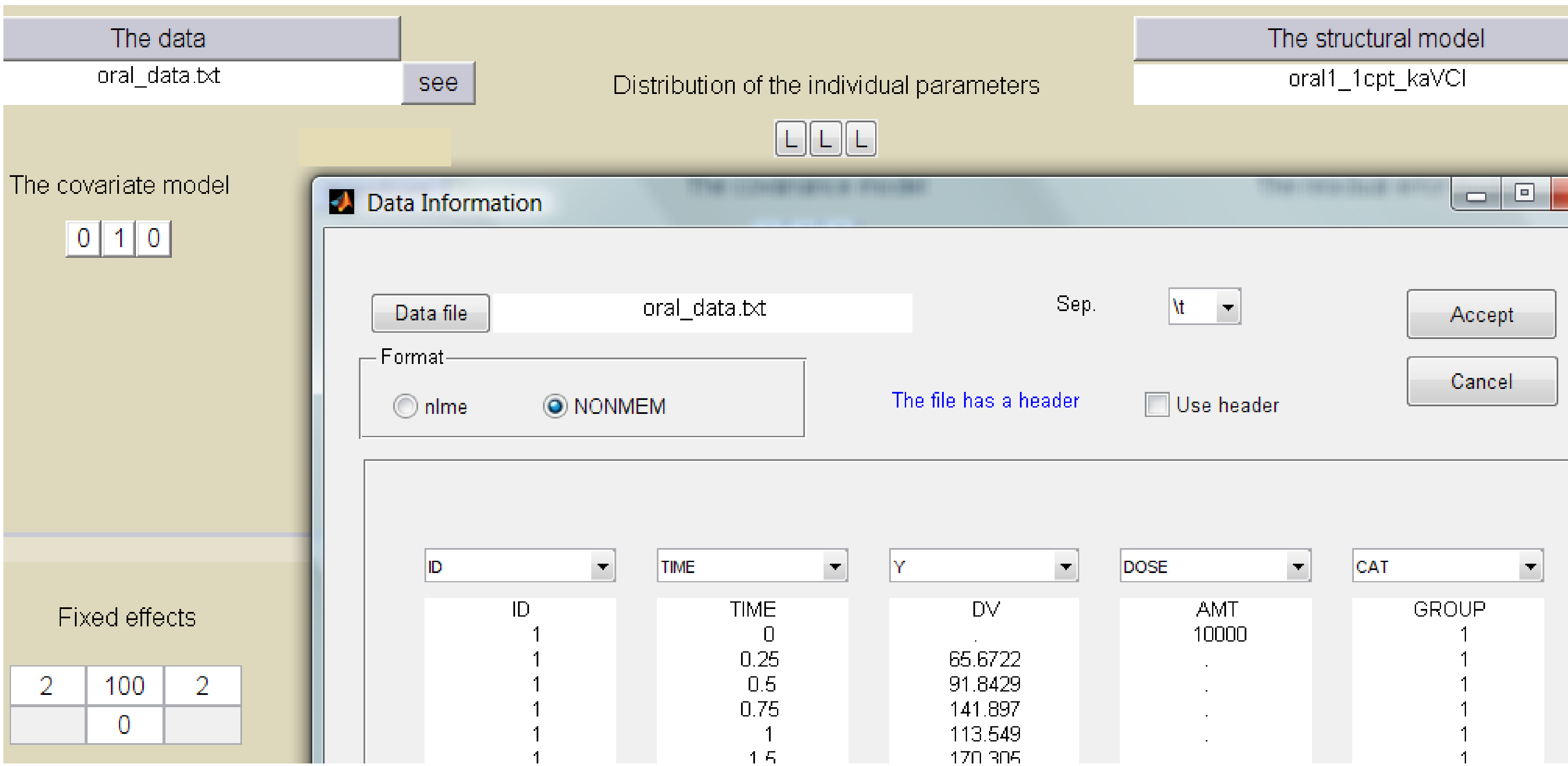
$$\text{Group 1} \quad \text{Vol}_i \sim \text{lognormal}(\mu_1, \omega_1)$$

$$\text{Group 2} \quad \text{Vol}_i \sim \text{lognormal}(\mu_2, \omega_2)$$

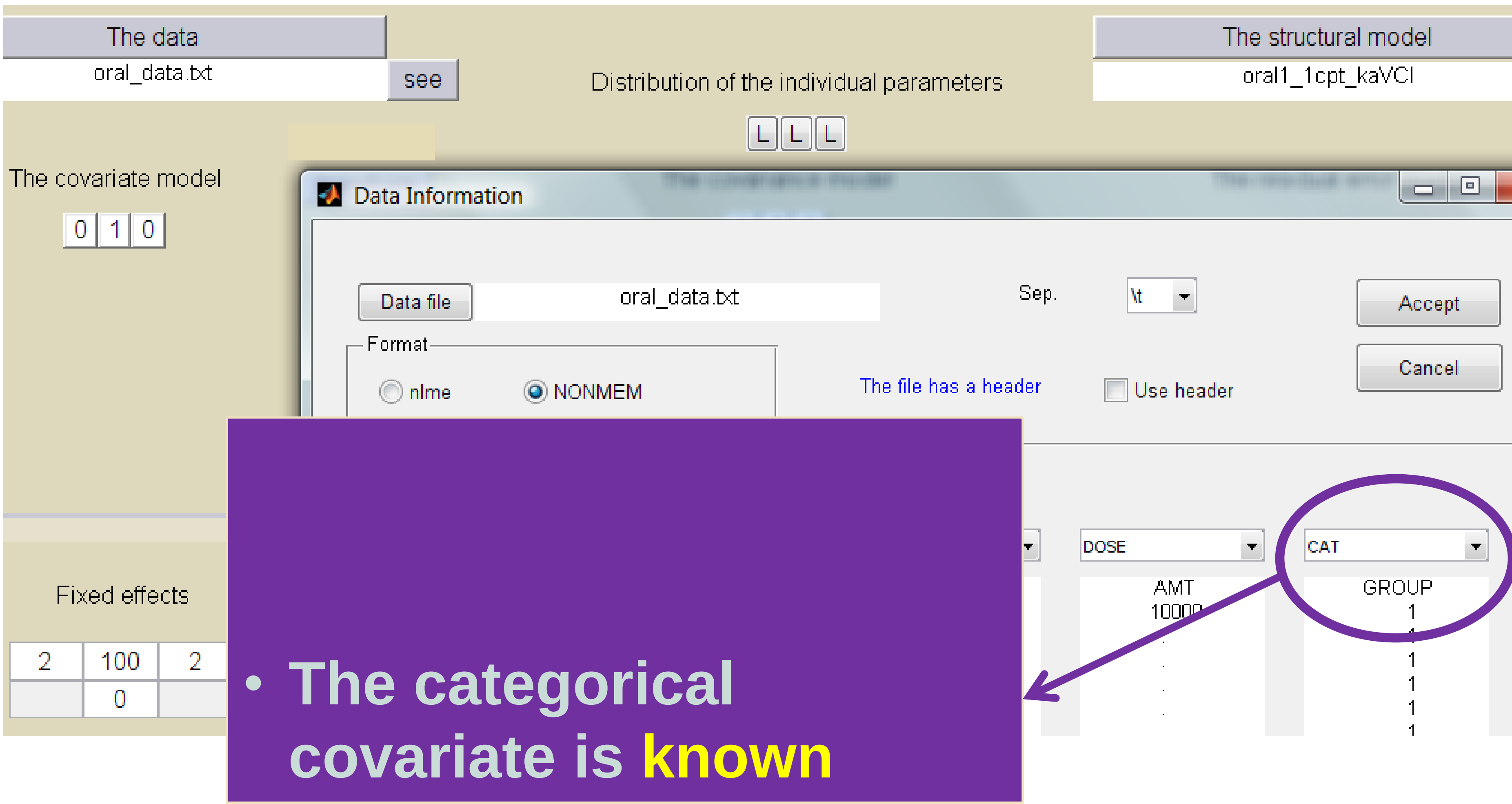


1) Supervised learning (the labels are known)

Categorical covariate model building with MONOLIX



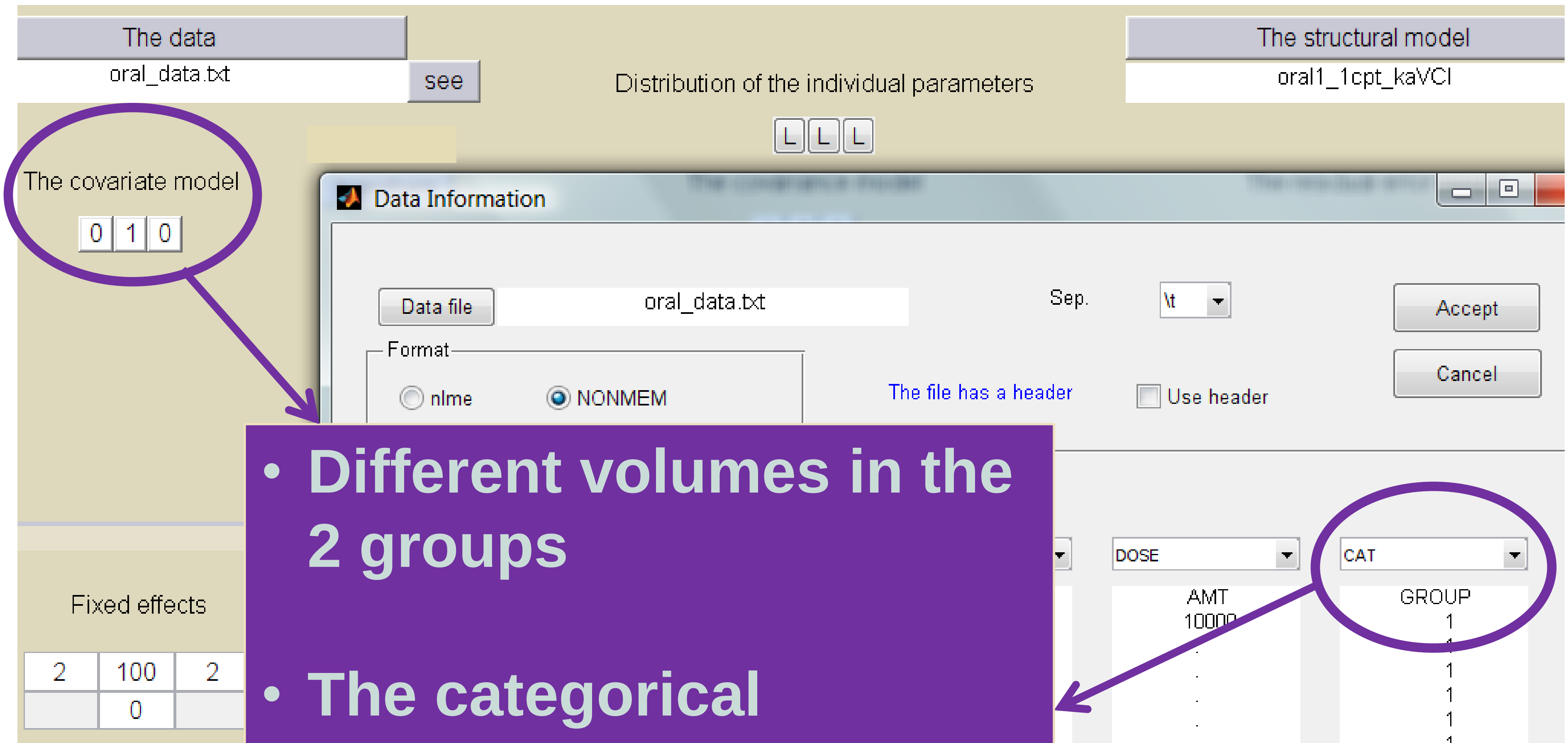
1) **Supervised learning** (the labels are known)
Categorical covariate model building with MONOLIX



- **The categorical covariate is known**

1) Supervised learning (the labels are known)

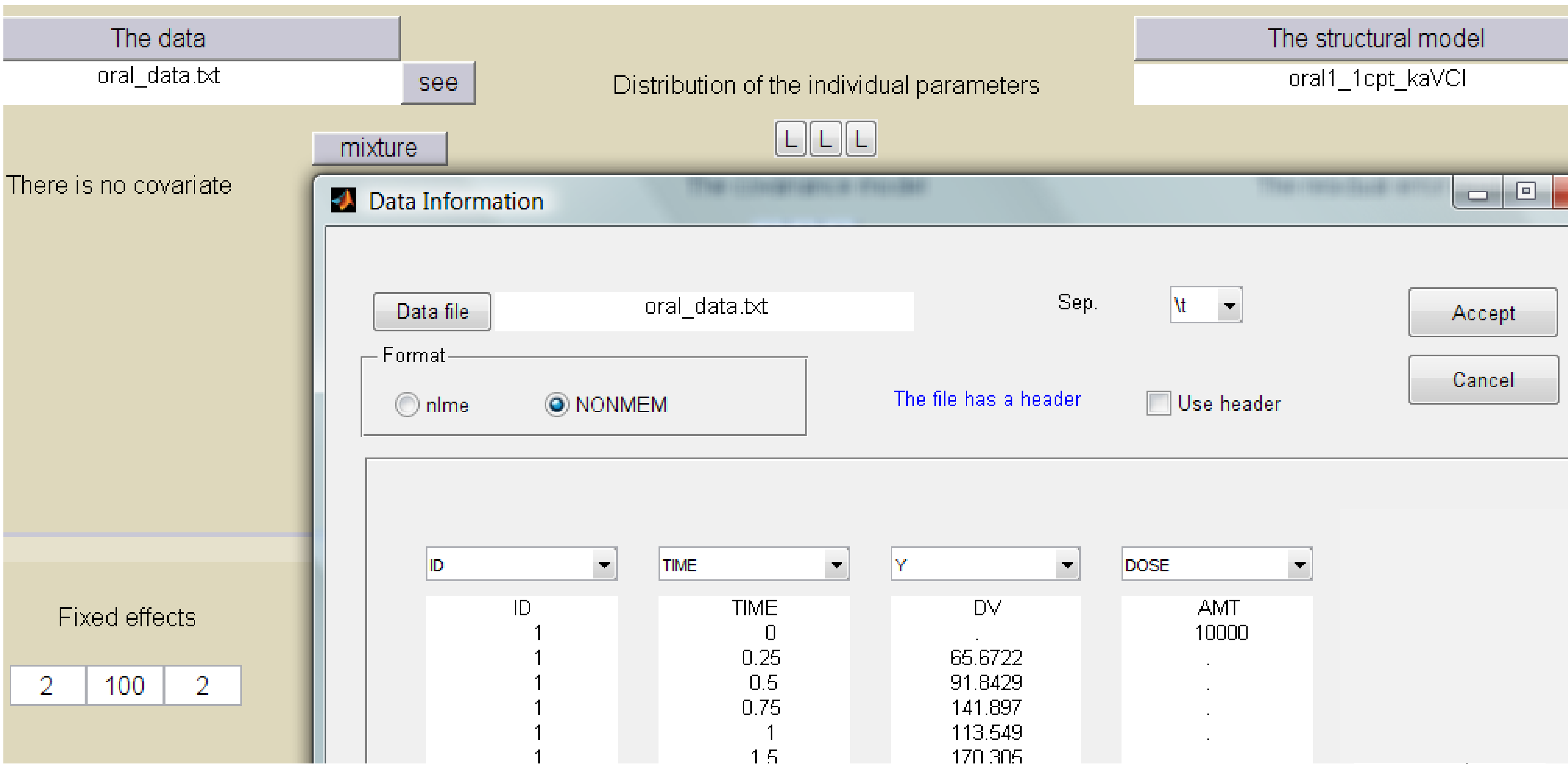
Categorical covariate model building with MONOLIX



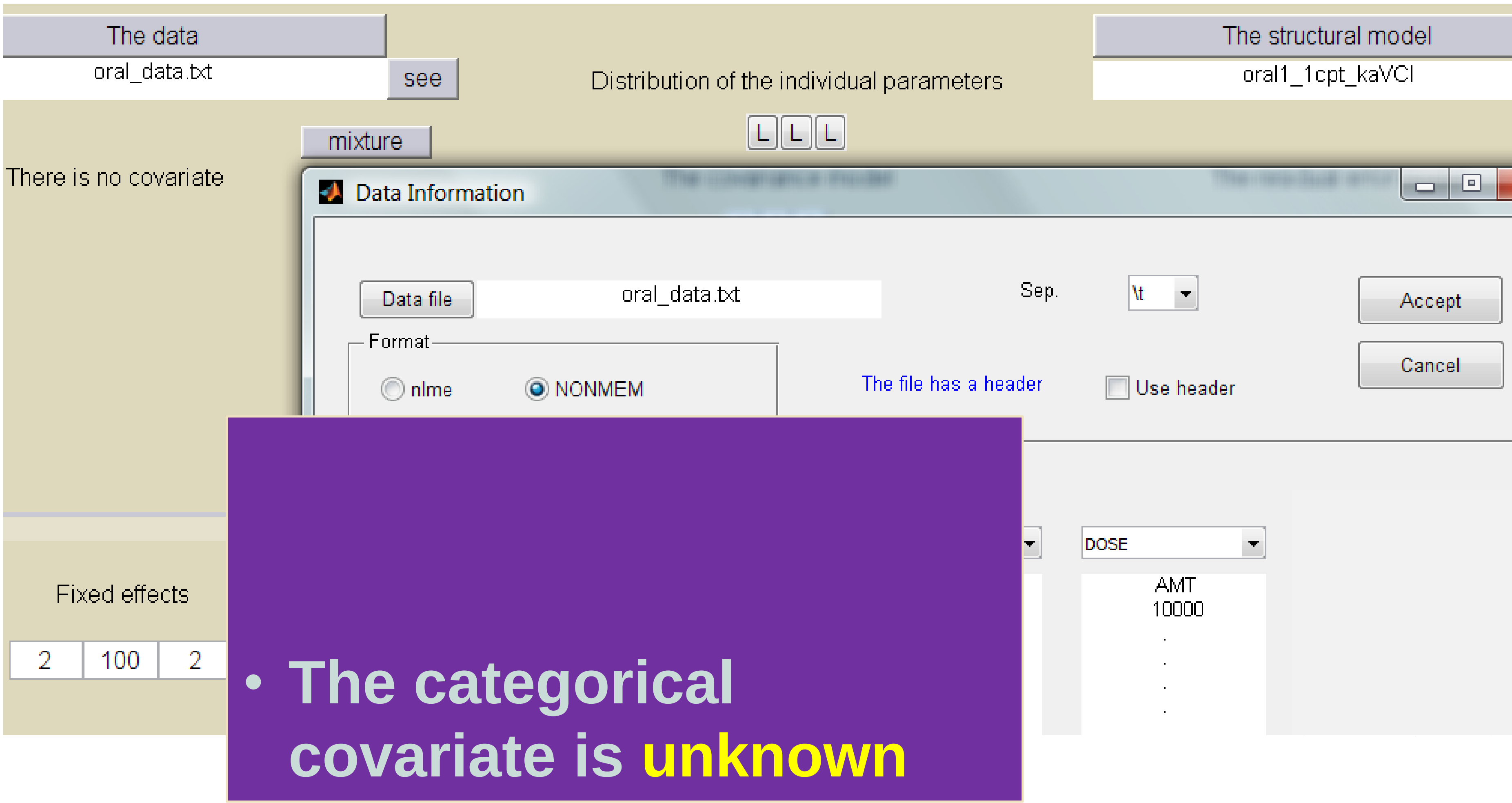
- **Different volumes in the 2 groups**
- **The categorical covariate is known**

2) Unsupervised learning (the labels are unknown)

Mixture model building with MONOLIX

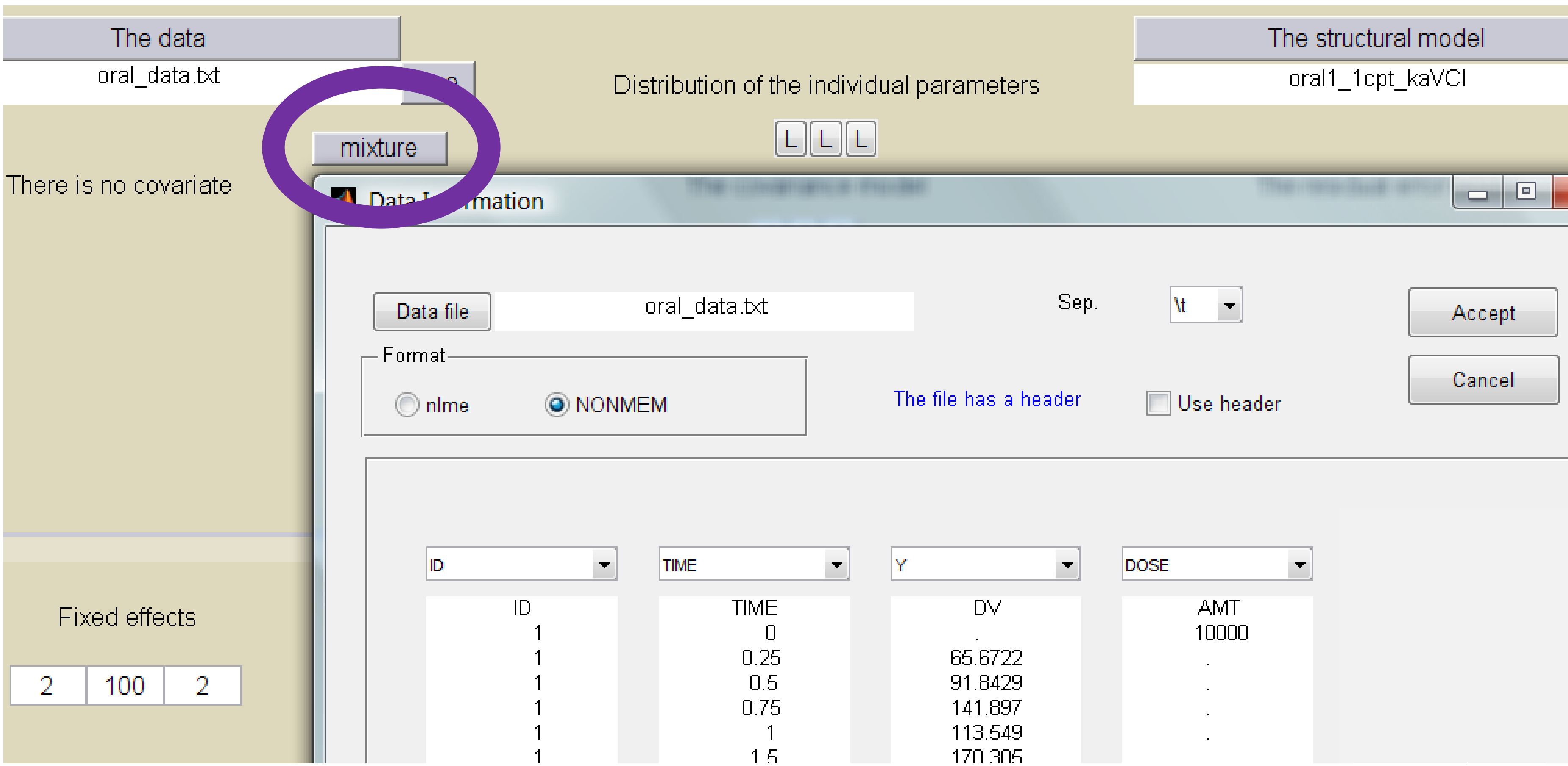


2) **Unsupervised learning** (the labels are unknown)
Mixture model building with MONOLIX

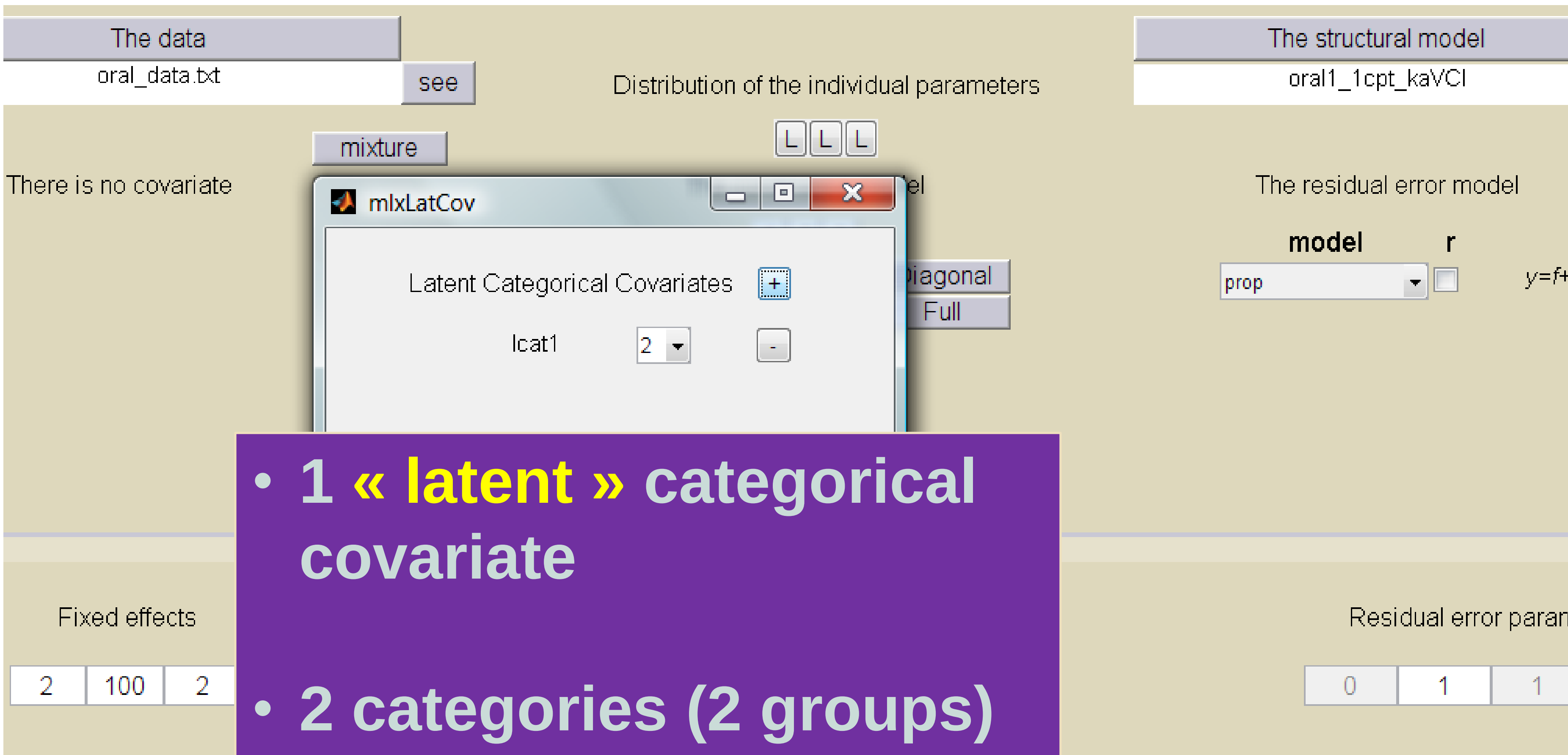


- The categorical covariate is **unknown**

2) Unsupervised learning (the labels are unknown)
Mixture model building with MONOLIX

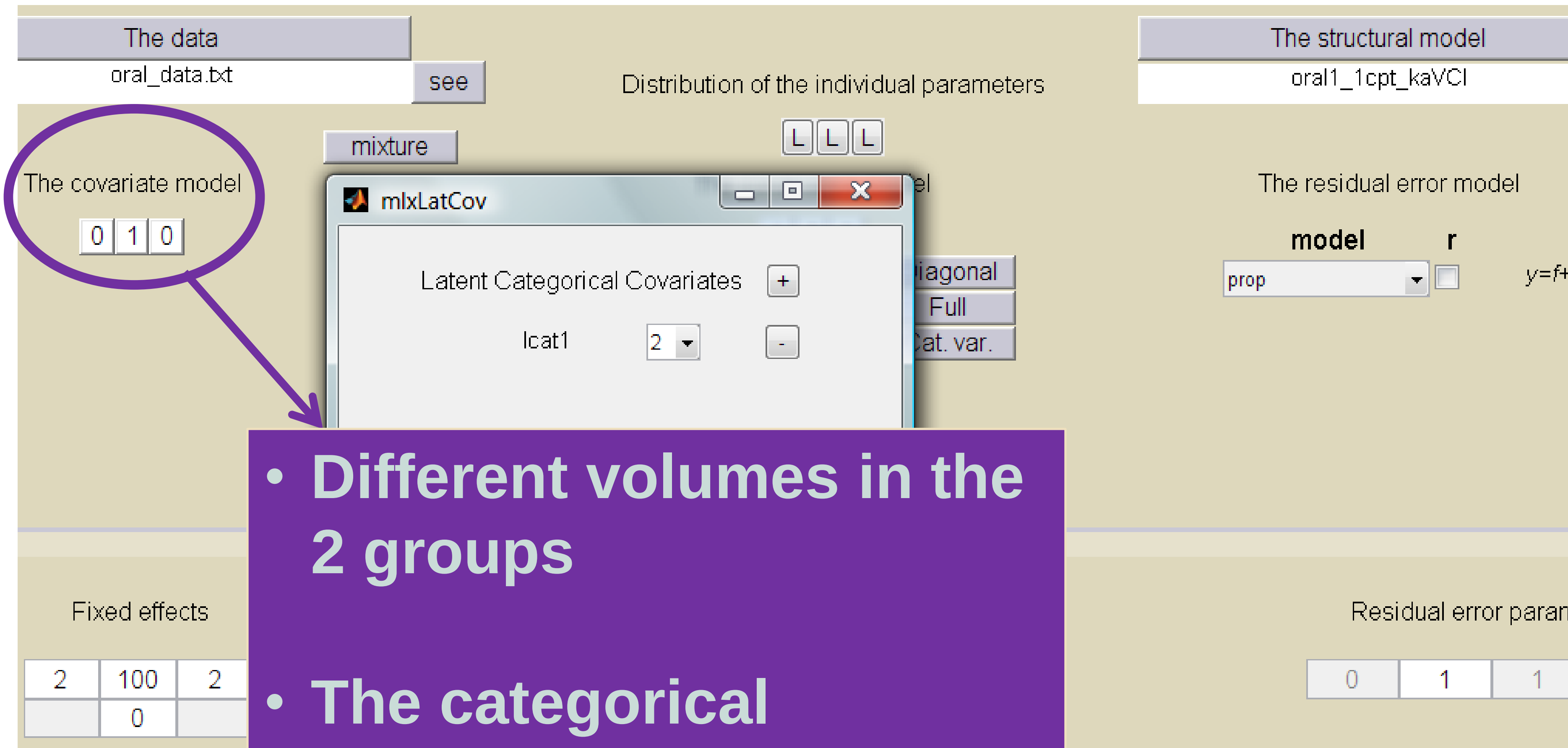


2) **Unsupervised learning** (the labels are unknown)
Mixture model building with MONOLIX



- 1 « **latent** » categorical covariate
- 2 categories (2 groups)

2) **Unsupervised learning** (the labels are unknown)
Mixture model building with MONOLIX

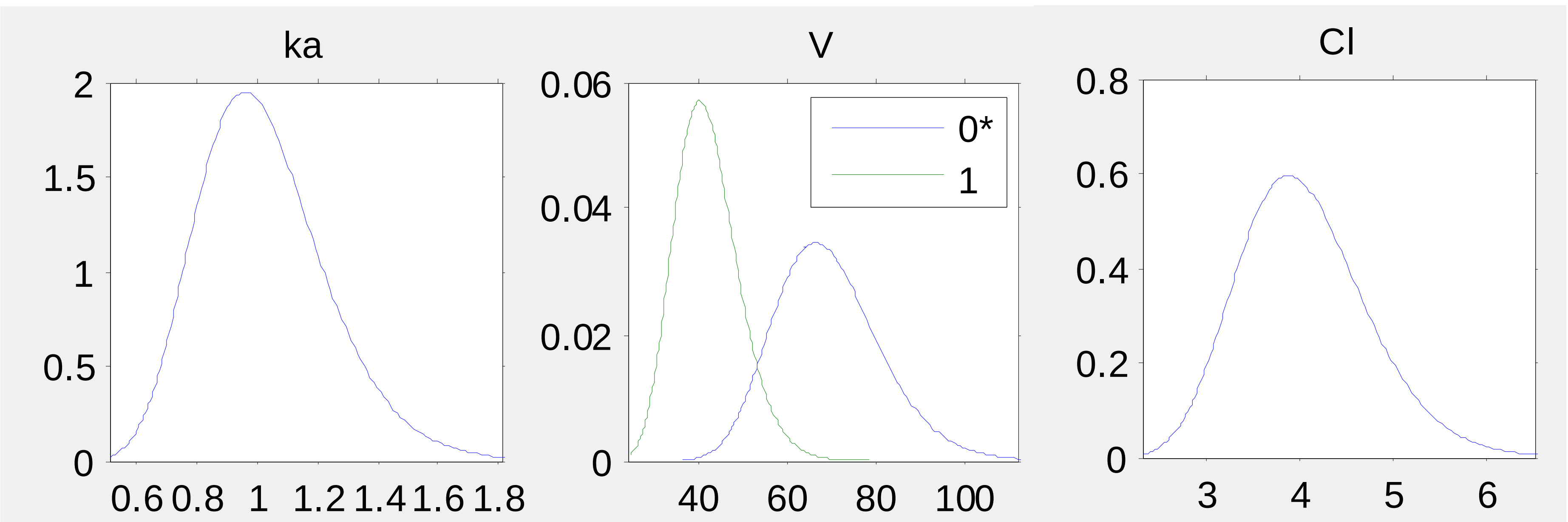


- **Different volumes in the 2 groups**
- **The categorical covariate is estimated**

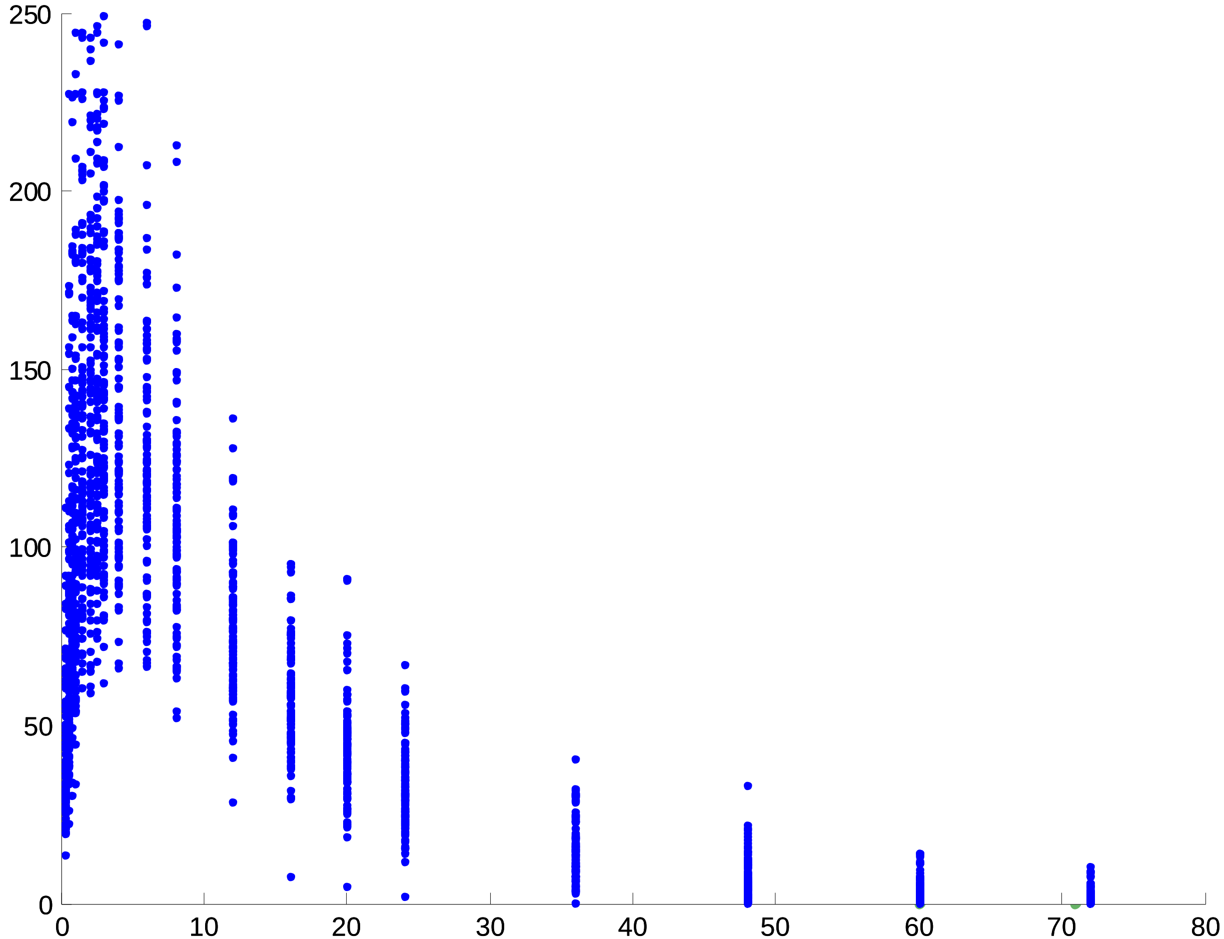
The estimated population parameters (SAEM for mixtures)

	« true values »	Known categorical covariate: estimations	Known categorical covariate: r.s.e.(%)	Estimated categorical covariate: estimations	Estimated categorical covariate: r.s.e.(%)
ka	1	1	3	1	3
V	70	68.3	2	65.6	3
β_V	-0.5	-0.507	8	-0.517	10
Cl	4	4	2	4	2
ω_{ka}	0.2	0.197	11	0.208	11
ω_V	0.2	0.179	8	0.186	11
ω_{Cl}	0.2	0.169	8	0.168	7
b	0.2	0.199	2	0.199	2
π_1	0.6	0.6	-	0.686	10
π_2	0.4	0.4	-	0.314	21

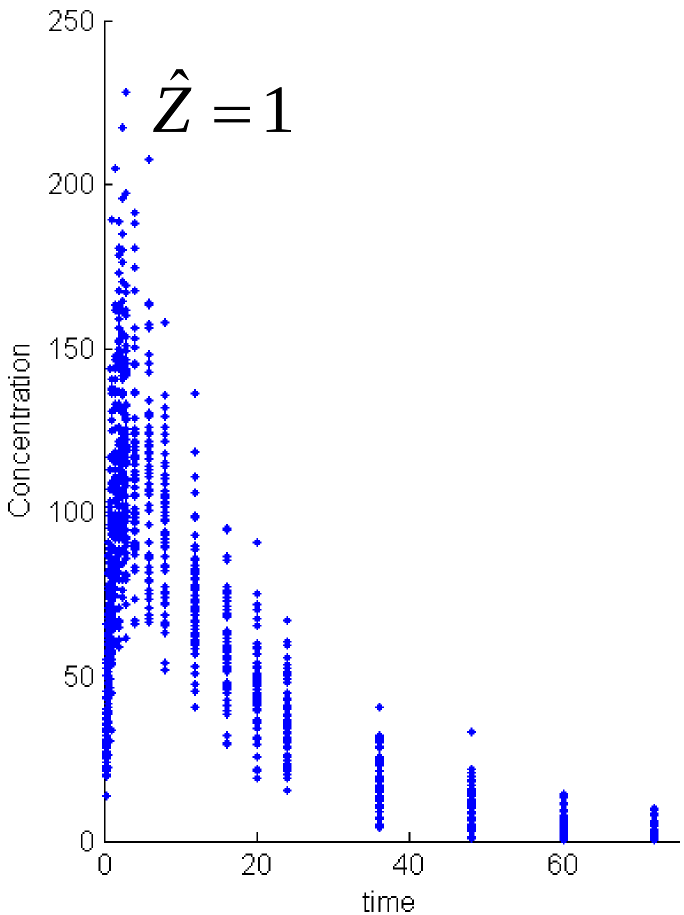
The estimated population distributions

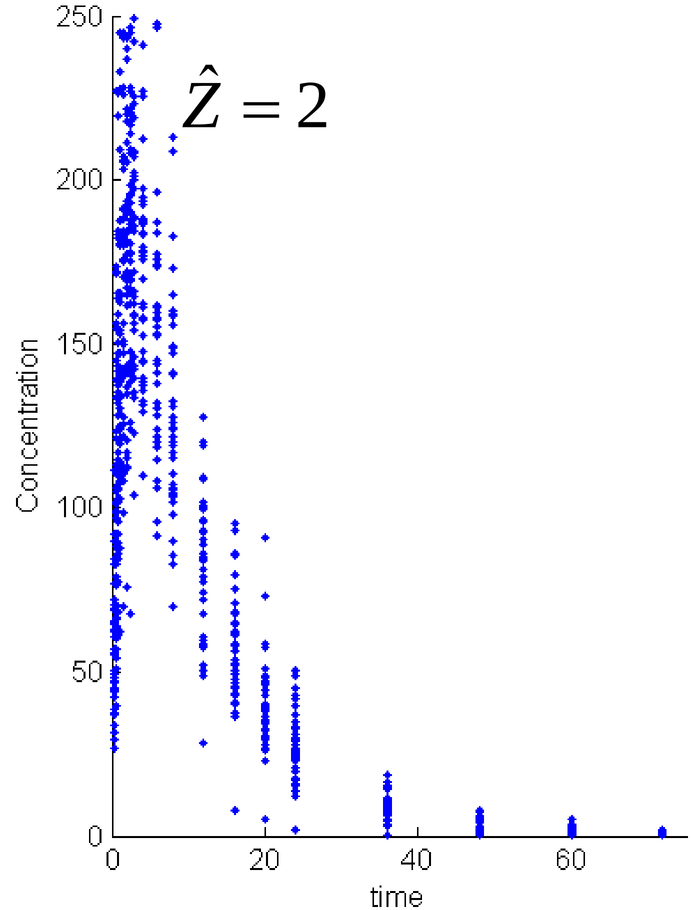


VPCs: global VPC includes both groups

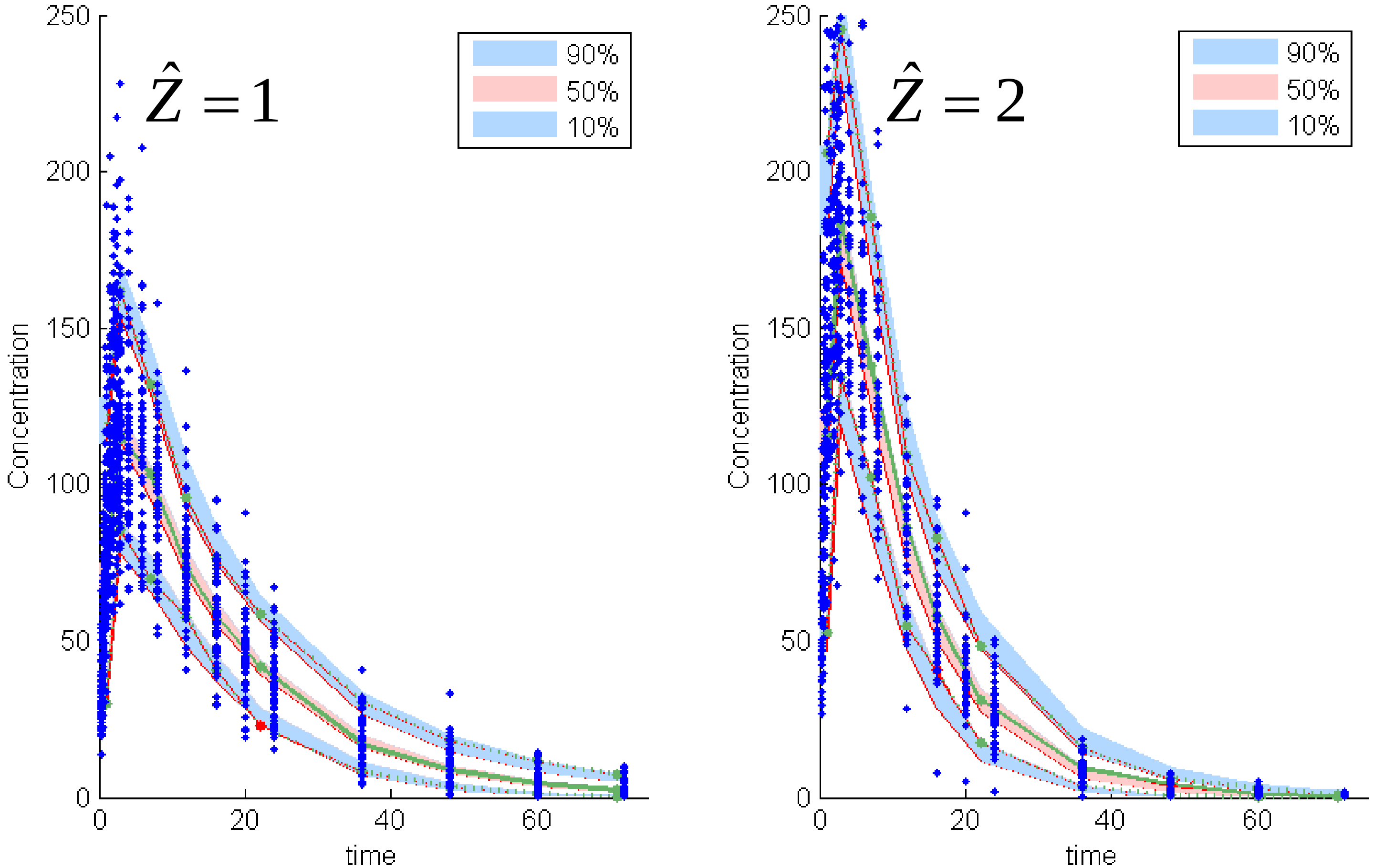


VPCs: use the estimated latent categorical covariate to stratify the data

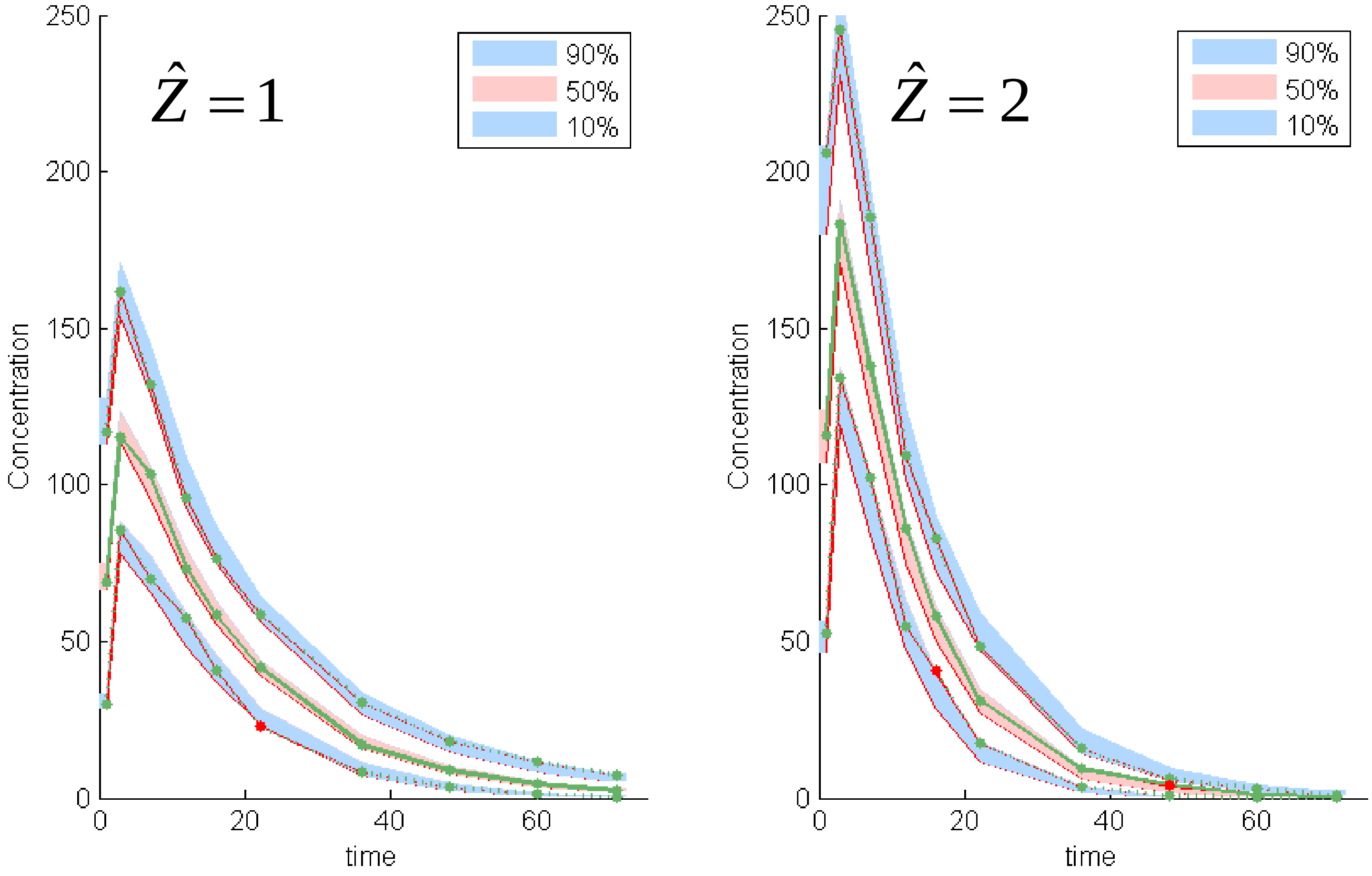




VPCs: use the estimated latent categorical covariate to stratify the data



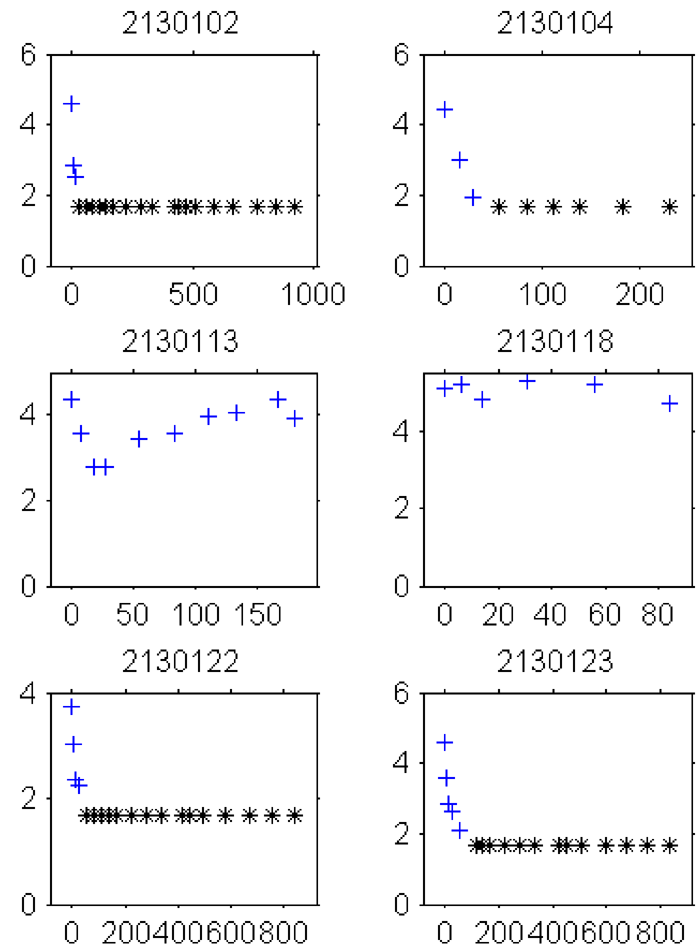
VPCs: use the estimated latent categorical covariate to stratify the data



II
Mixture of models

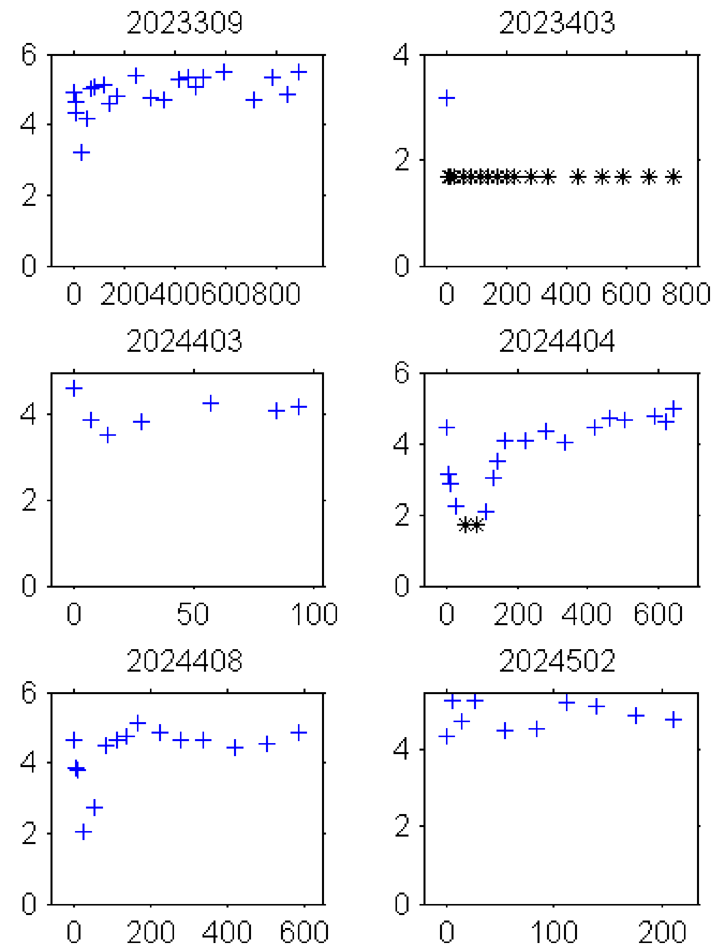
Some data

- POWER studies were conducted by TIBOTEC
- Viral load data from 578 HIV infected patients
- Randomized, controlled, partially blinded studies
- 3 studies of up to 144 weeks, performed in highly treatment experienced patients, using darunavir/ritonavir (DRV/RTV) or an investigator-selected control PI, combined with an optimised background regimen (OBR), consisting of nucleoside reverse transcriptase inhibitors with or without the fusion inhibitor enfuvirtude.



Some data

- POWER studies were conducted by TIBOTEC
- Viral load data from 578 HIV infected patients
- Randomized, controlled, partially blinded studies
- 3 studies of up to 144 weeks, performed in highly treatment experienced patients, using darunavir/ritonavir (DRV/RTV) or an investigator-selected control PI, combined with an optimised background regimen (OBR), consisting of nucleoside reverse transcriptase inhibitors with or without the fusion inhibitor enfuvirtude.

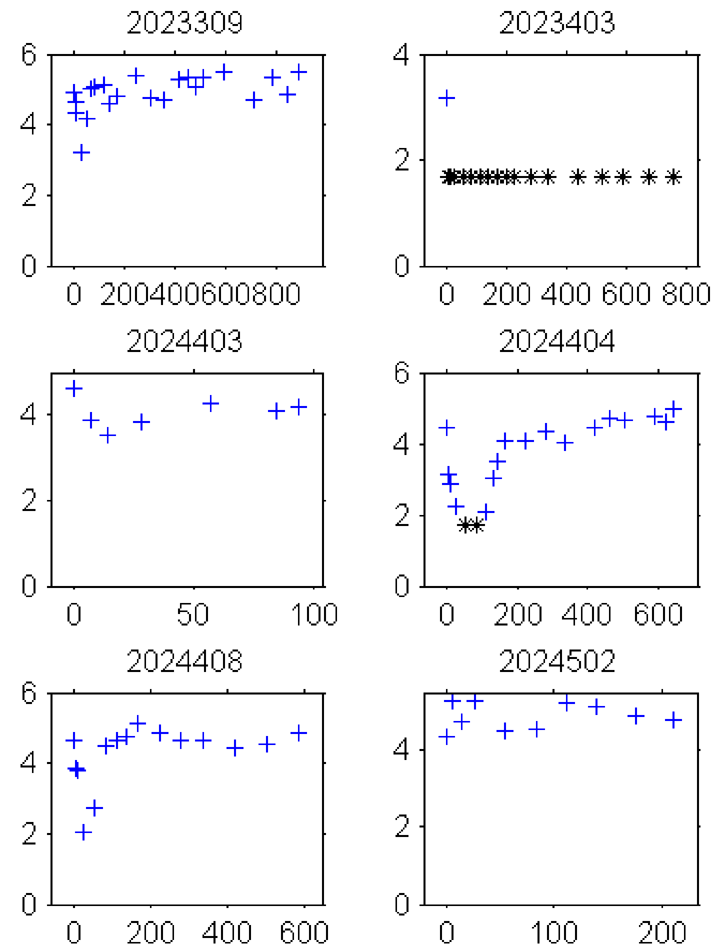


Some data

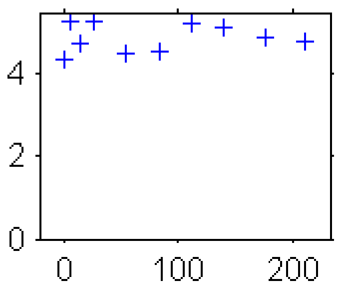
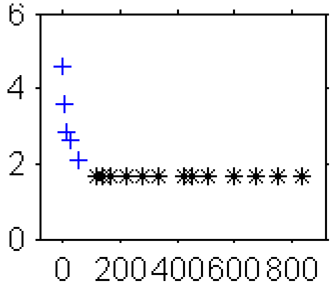
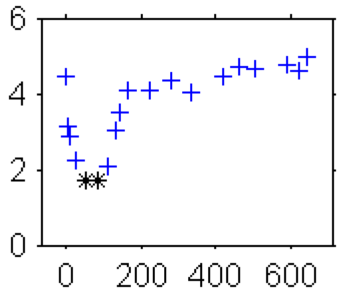
The data exhibit three different typical profiles:

- non-responders,
- responders,
- rebounders.

=> We propose to describe these viral load data with a mixture of three models



Between subject model mixture
(unsupervised learning)

3 different profiles	3 different groups	3 different VK models
	nonresponder $z_i = 1$	$f_1(t) = A_1 + A_2$
	responder $z_i = 2$	$f_2(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}$
	rebounder $z_i = 3$	$f_3(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + \dfrac{A_3}{1 + e^{-\lambda_3 (t - t^*)}}$

Between subject model mixture
(unsupervised learning)

$$f_1(t) = A_1 + A_2$$

$$f_2(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}$$

$$f_3(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + \frac{A_3}{1 + e^{-\lambda_3 (t - t^*)}}$$

$$p_1 = P(z_i = 1)$$

$$p_2 = P(z_i = 2)$$

$$p_3 = P(z_i = 3)$$

Population approach:

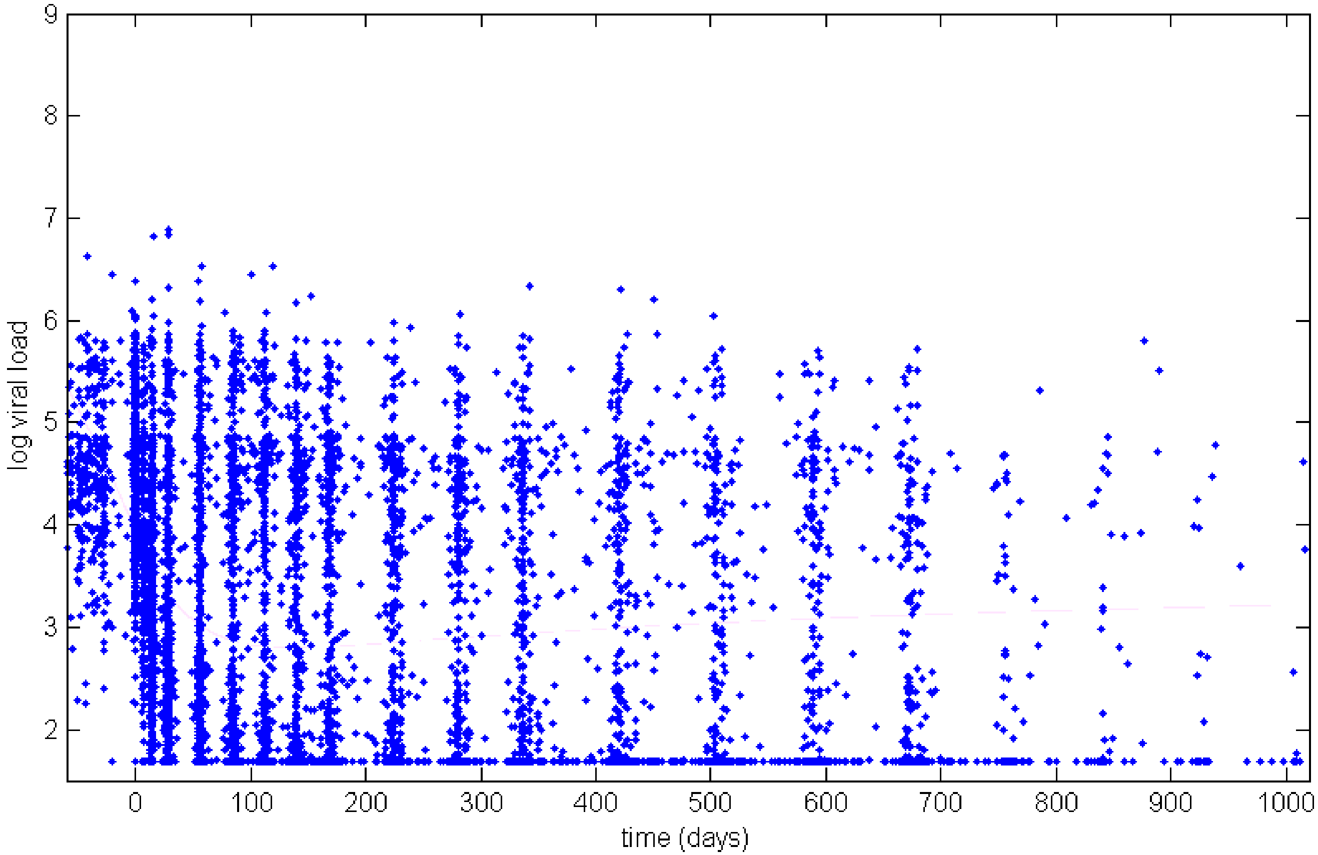
- IIV on A_1 , λ_1 , A_2 , λ_2 , A_3, λ_3 , t^*
- No variability on p_1 , p_2 , p_3

Between subject model mixture
MLXTRAN implementation

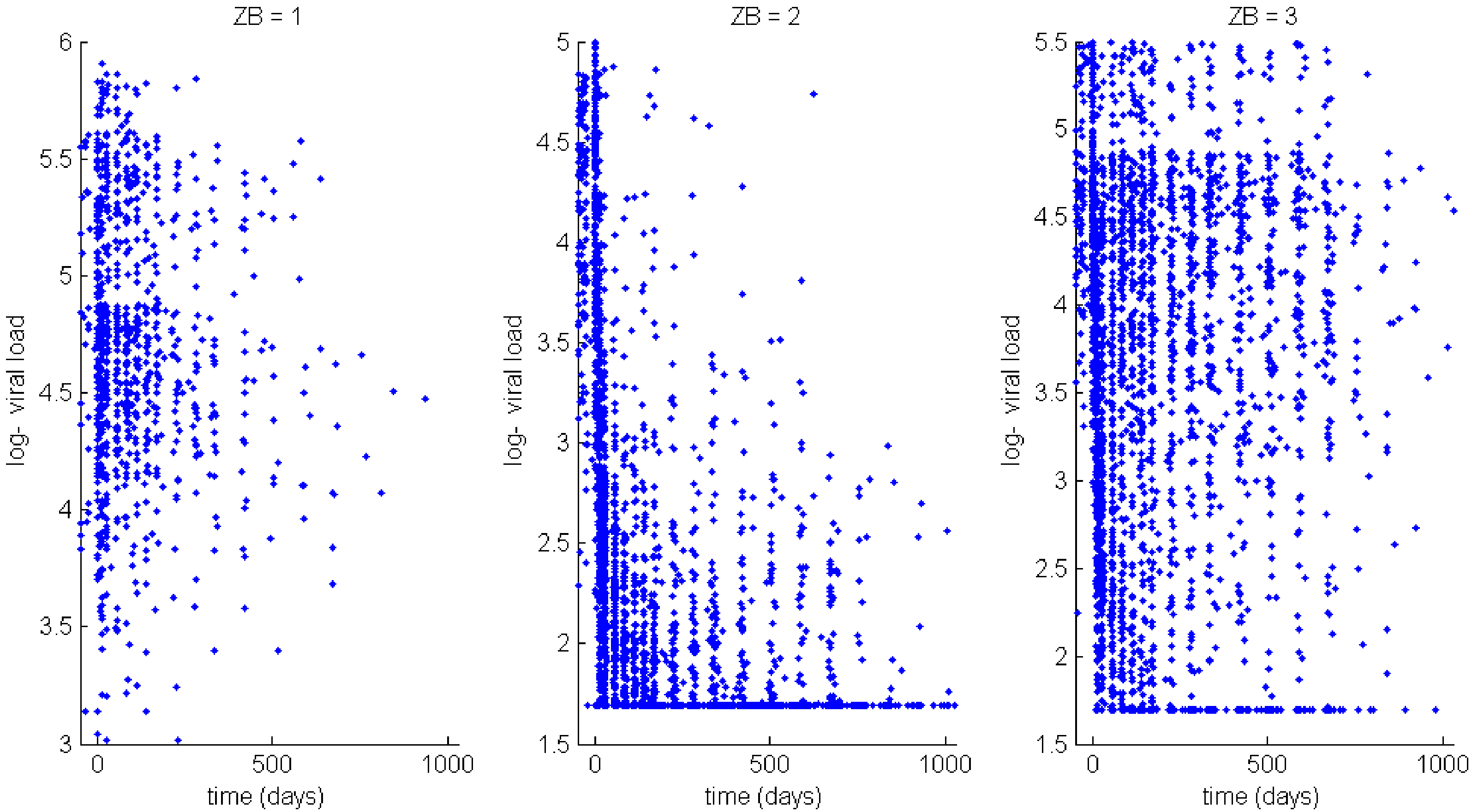
```
$PROBLEM    Between Subject Model Mixture


$PSI   A1 L1 A2 L2 A3 L3 TL S1 S2


$EQUATION
f1=A1+A2
f2=A1*exp(-L1*T)+A2*exp(-L2*T)
f3=A1*exp(-L1*T)+A2*exp(-L2*T)+A3/(1+exp(-L3*(T-TL)))
p1=1/(1+S1+S2)
p2=S1/(1+S1+S2)
p3=S2/(1+S1+S2)


$OUTPUT
OUTPUT1 = BSMM(f1,p1,f2,p2,f3,p3)
```

Between subject model mixture

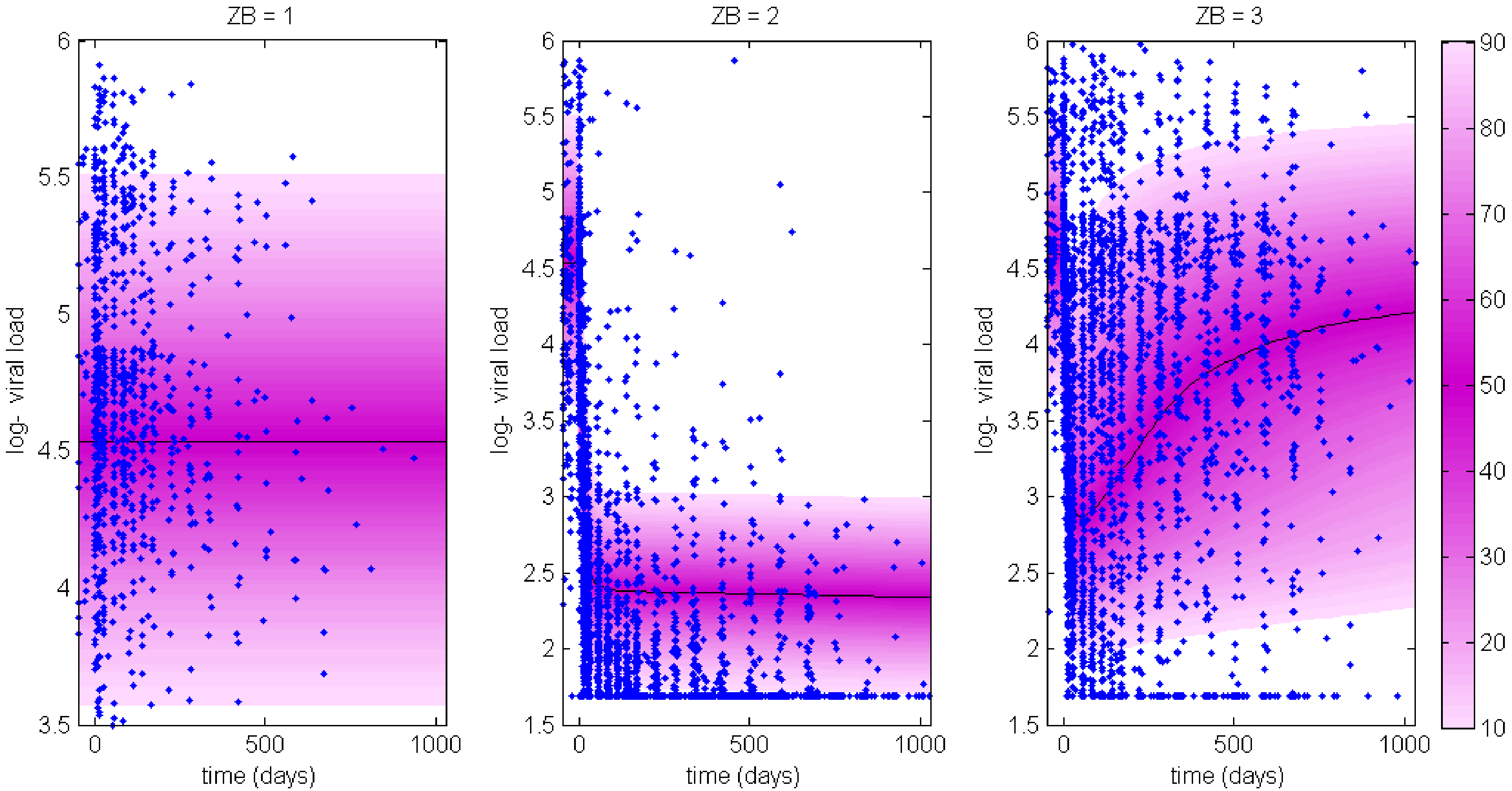
Some results



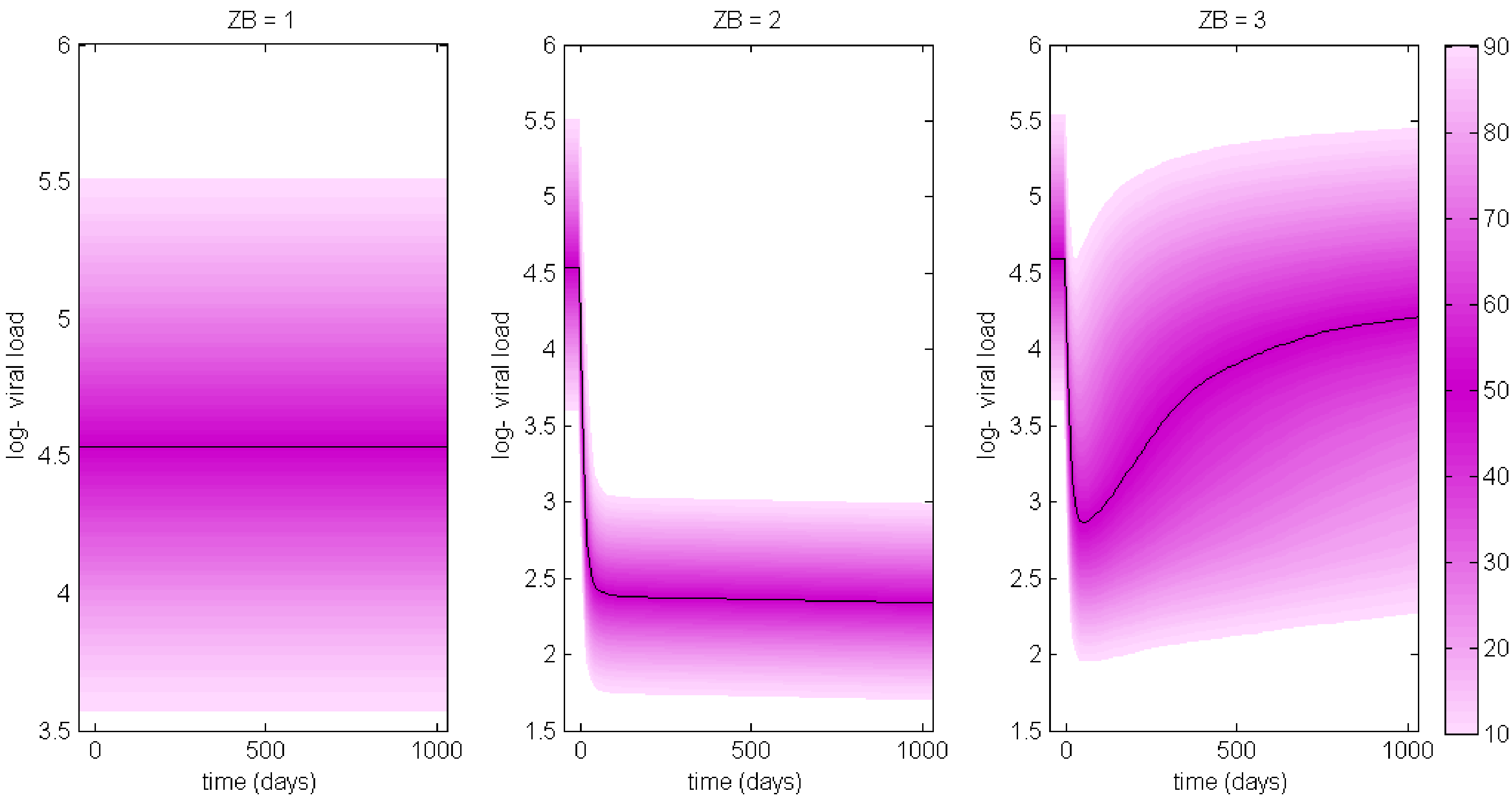
Between subject model mixture
Some results



Between subject model mixture
Some results

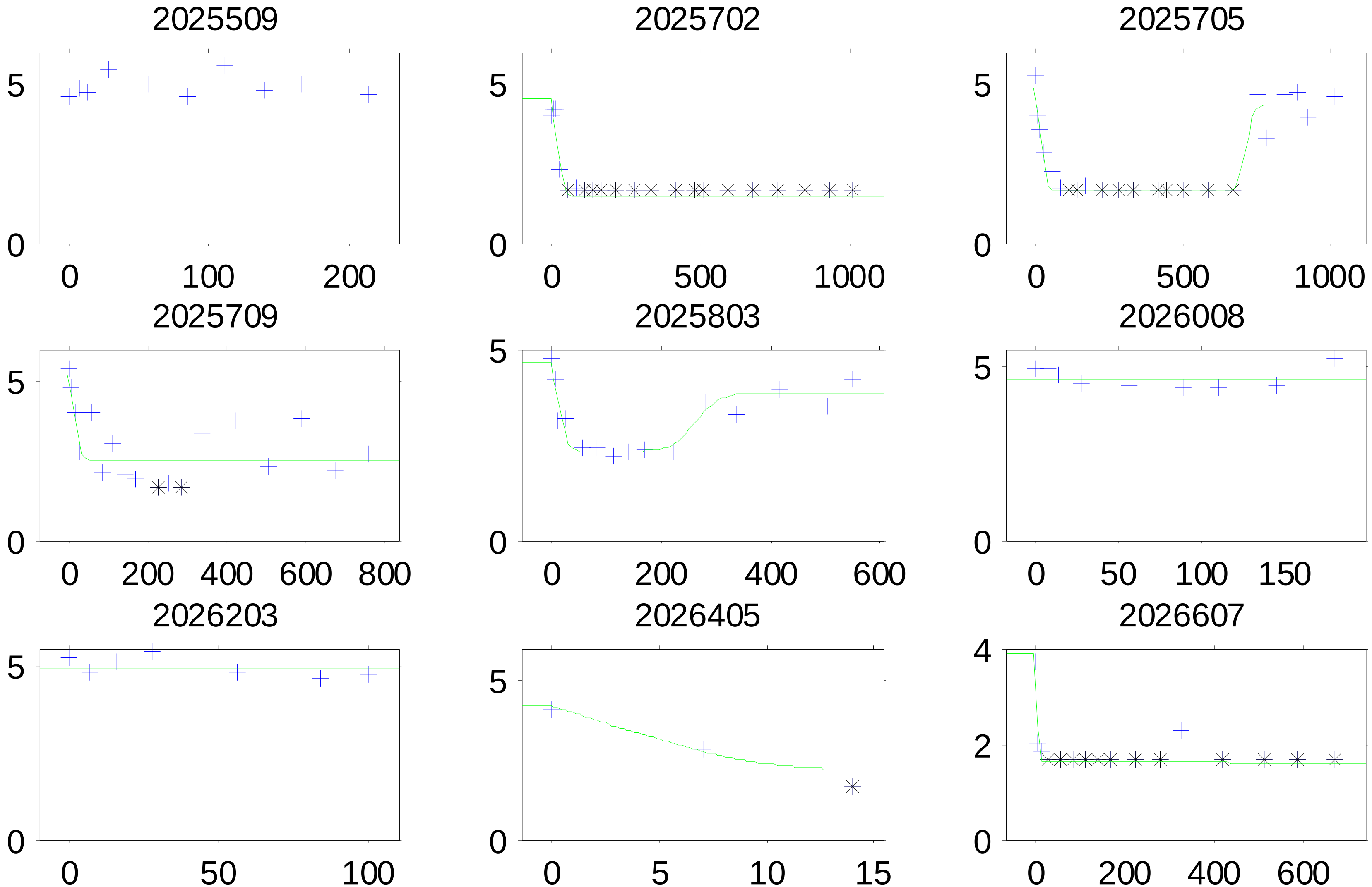


Between subject model mixture
Some results



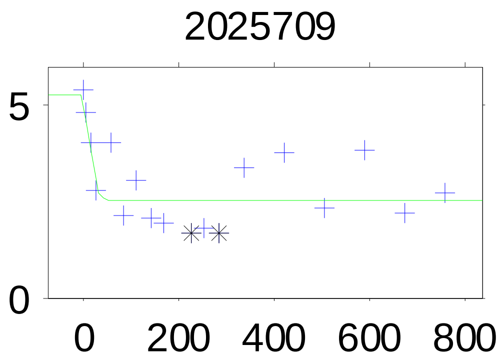
Between subject model mixture

Some results



Between subject model mixture

Some results



Should we consider this subject

- as a non responder?
- as a responder?
- as a rebounder?

Or someone « in between »?

Within subject model mixture

$$\left.\begin{aligned} f_1(t) &= A_1 + A_2 \\ f_2(t) &= A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} \\ f_3(t) &= A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + \frac{A_3}{1+e^{-\lambda_3 (t-t^*)}} \end{aligned}\right\}$$

$$\boxed{f = p_1 f_1 + p_2 f_2 + p_3 f_3}$$

Population approach:

- IIV on A_1, λ_1, A_2, λ_2, A_3, λ_3, t^*
- **IIV on p_1, p_2, p_3**

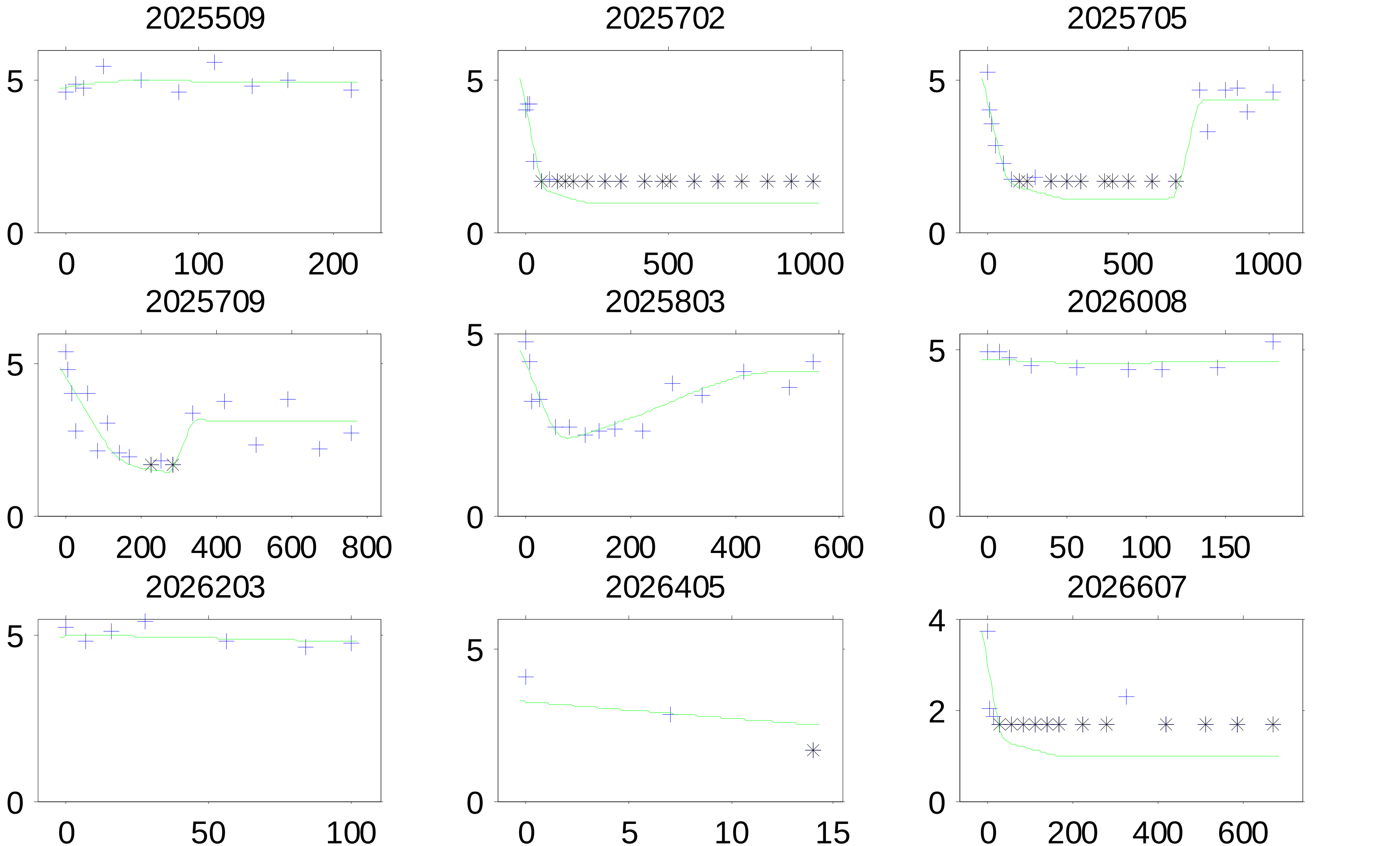
MONOLIX IMPLEMENTATION

Within Subject Model Mixture

```
$PROBLEM    Within  Subject Model Mixture


$PSI   A1  L1  A2  L2  A3  L3  TL  S1  S2


$EQUATION
f1=A1+A2
f2=A1*exp(-L1*T)+A2*exp(-L2*T)
f3=A1*exp(-L1*T)+A2*exp(-L2*T)+A3/(1+exp(-L3*(T-TL)))
p1=1/(1+S1+S2)
p2=S1/(1+S1+S2)
p3=S2/(1+S1+S2)


$OUTPUT
OUTPUT1 = WSMM(f1,p1,f2,p2,f3,p3)
```

Within subject model mixture

Some results



SAEM for mixture models

at iteration k,

i) **simulate** the latent categorical covariates *(z_i)* and the individual parameters *(ψ_i)*

$$\left(z_i^{(k)},\psi_i^{(k)}\right)\sim p\left(z_i,\psi_i \middle| y_i;\theta^{(k)}\right)$$

ii) **estimate** the expectation of the complete log-likelihood using stochastic approximation

$$Q_{k+1}(\theta)=Q_k(\theta)+\gamma_k\left(LL\left(y_i,z_i^{(k)},\psi_i^{(k)};\theta\right)-Q_k(\theta)\right)$$

iii) **update** the estimation of the population parameters

$$\theta_{k+1}=\operatorname{Arg}\max Q_{k+1}(\theta)$$

Conclusions

- The SAEM algorithm was extended for the analysis of mixture models
- The algorithm handles different types of mixtures (mixture distributions, between and within model mixtures)
- The estimated labels can be used to stratify the data
- The algorithms are implemented in new MONOLIX 3.2 and supported by MLXTRAN
- Other possible extensions are straightforward (mixture of error models...)