Mixture Modelling for the Detection of Subpopulations

A. Lemenuel-Diot, N. Frey, C. Laveille,
R. Jochemsen and A. Mallet

Problem

Are we really sure of the hypothesis of homogeneity in a population ?

Is it possible to find heterogeneities without assumption about

-their existence ?

-their origin ?

Method

Mixture models & Regression methods

1) Detection of Heterogeneity

2) Explanation of Heterogeneity

Likelihood Maximization & Gauss-Hermite Quadrature

3) Estimation fixed parameters and random parameters distribution

Definition of the likelihood

Likelihood formula : $L=\Pi\ L_i$

with $L_i = \int_{R^n} g_i(y_i,\theta,Z_i,\eta).h(\eta).d\eta$

i: subject
$\mathbf{y}_i$: observations
$\mathbf{Z}_i$: covariates
θ: fixed parameters
η: random parameters
g_i: distribution of the observations
h : distribution of the random effect

Likelihood Approximation

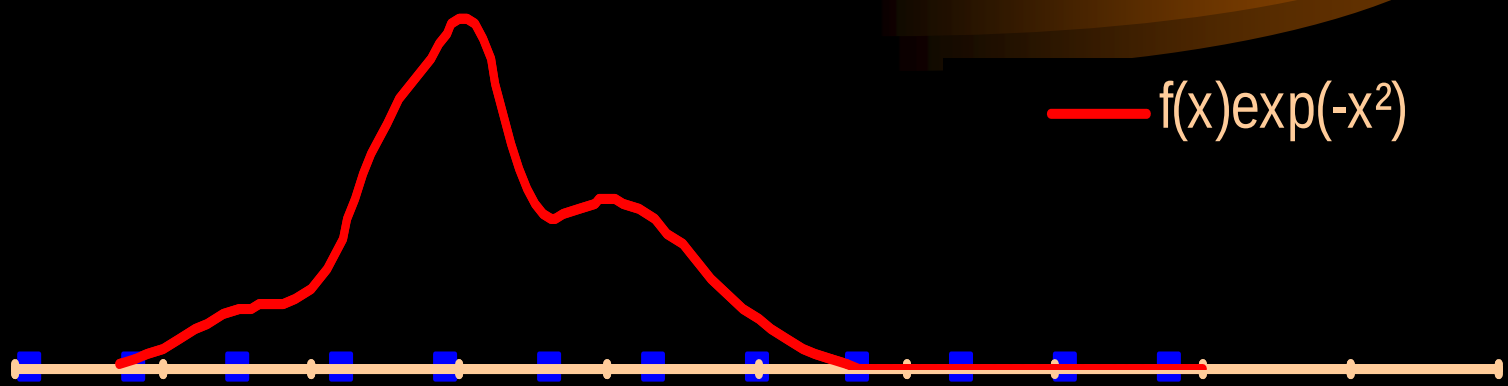
Different possible methods

Taylor approximation (NONMEM)

Non parametric approach NPML

Gauss-Hermite Quadrature

Gauss-Hermite Quadrature



$$\int_R f(x).e^{-x^2}dx = \Sigma_{k=1,\ldots,r} f(x_k).A_k$$

r: order of the quadrature

A_k: weights of the quadrature

x_k: nodes of the quadrature

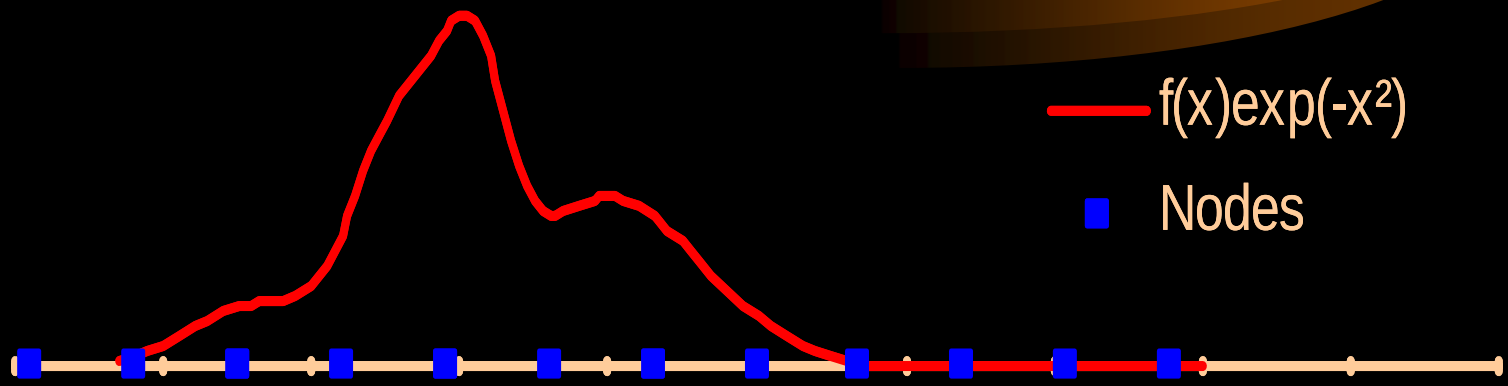
Use of GH quadrature to approximate L

$$L_i = \int_{R^n} g_i(y_i,\theta,Z_i,\eta).h(\eta).d\eta$$

$$\int_R f(x).e^{-x^2}dx = \Sigma_{k=1,\ldots,r} f(x_k).A_k$$

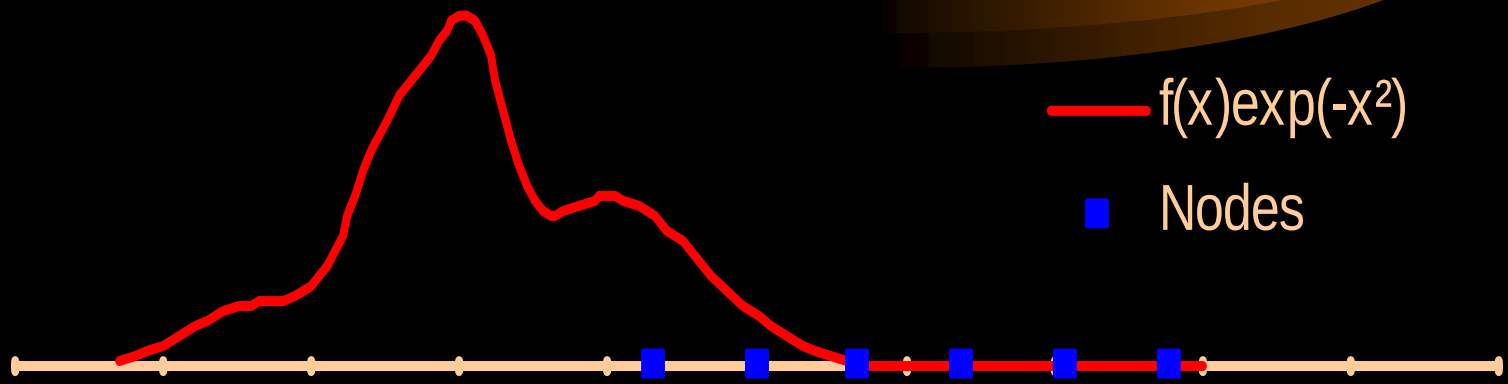
Advantage : Easily identifiable to the likelihood function

Use of GH quadrature to approximate L



Advantage : Quality of approximation increases with the order

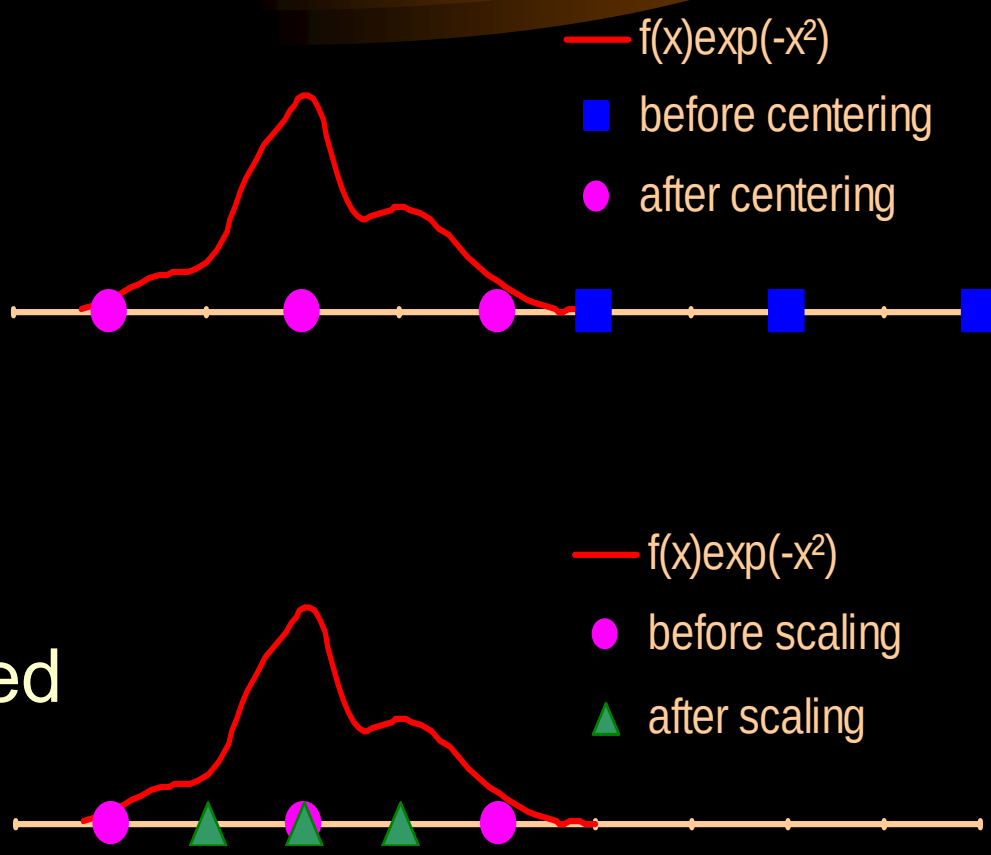
Use of GH quadrature to approximate L



Drawback : Nodes fixed whatever the function f (found in GH table)

Adaptation of the Quadrature Nodes

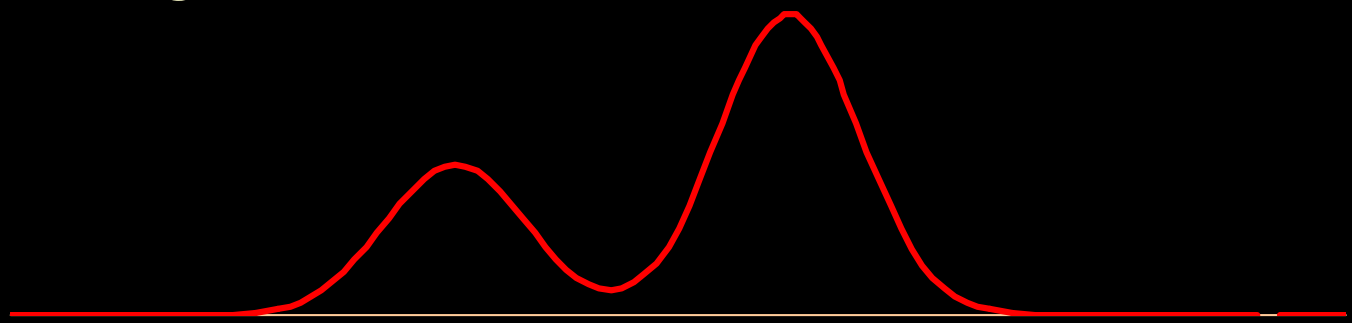
- Centering
- Scaling optimized



Detection of Heterogeneity

Mixture Model

Model in which the joint distribution of random effects is assumed to be a weighted sum of normal distributions



Test for the Number of Subpopulations

Γ_p : Model p components / Γ_q : Model q components

Kullback-Leibler test

p<q , Null Hypothesis : Model with p components

$$2LR=2\Sigma_{i=1,\ldots,N} \log(L_{i\,\Gamma_q} / L_{i\,\Gamma_p})$$

Asymptotic distribution of $2LR$ under the null hypothesis → Weighted sum of chi2 1df

Attribution of patients subpopulations G_i

Most Probable Subpopulation G_i for subject i

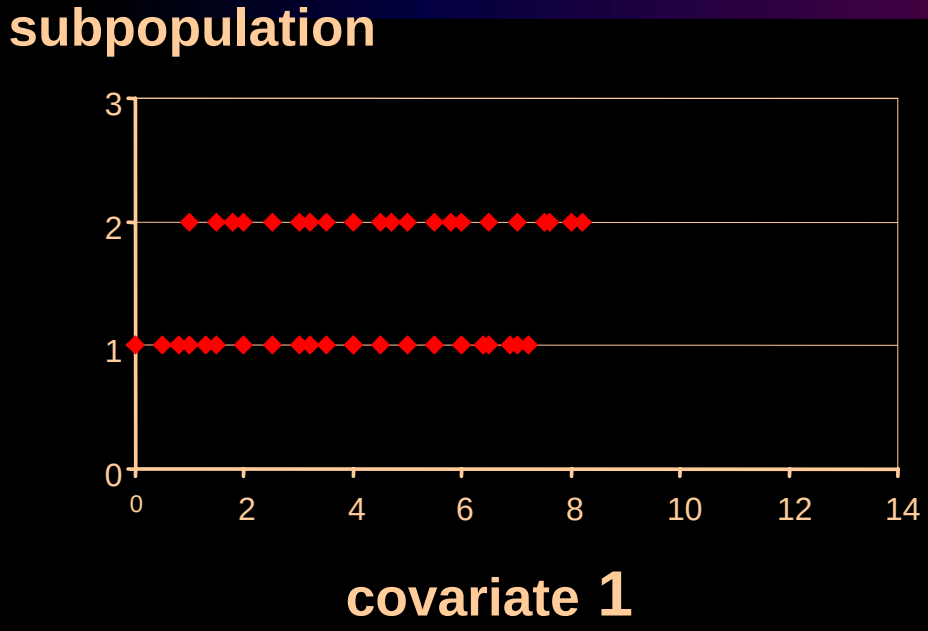
supposing p subpopulations and j=1,…,p

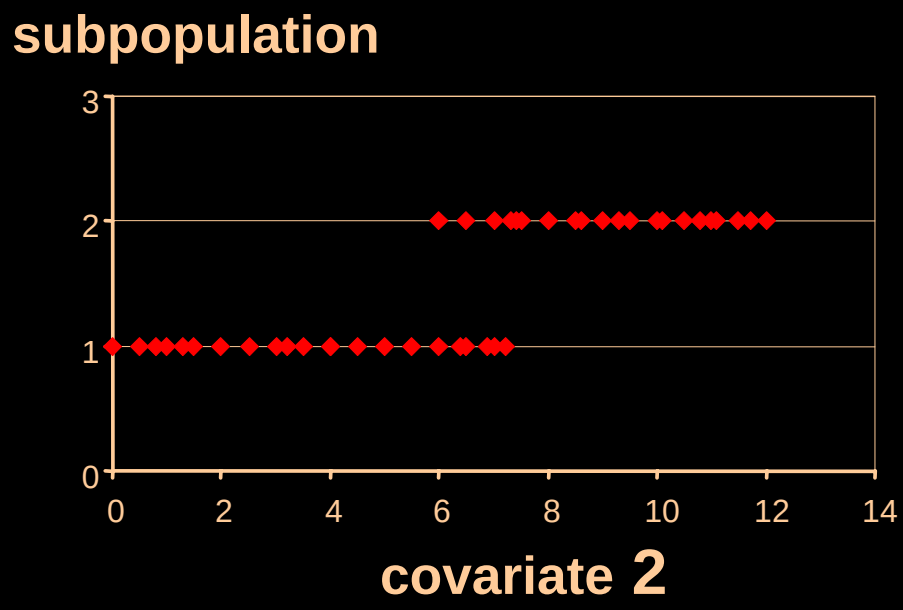
$$\mathrm{Max}_j\,[P(G_i=j)]$$

with $P(G_i=j) = \alpha_j L_{ij} / \Sigma_{k=1,\ldots,p}\, \alpha_k L_{ik}$ and

L_{ij} the likelihood associated to the j^{th} subpopulation

Relationship subpopulations-covariates





Relationship covariates-parameters

m Classes of Covariates significant

Non-Adapted Model

$$\text{Parameter}(Z)=\theta.\exp(\eta)$$

Adapted Model

$$\text{Parameter}(Z)= [\Sigma_k \theta_k 1_{Z=Z_k}].\exp(\eta)$$

Analysis strategy

Global Analysis

Same number of subpopulations on the distribution of all parameters

Analysis parameter by parameter

Only one parameter has many subpopulations

Application : S05702

Gliclazide : antidiabetic, treatment of type II diabetes

Data : collected during the clinical development of a once a day modified release formulation

2 phase II + 2 phase III → 634 patients

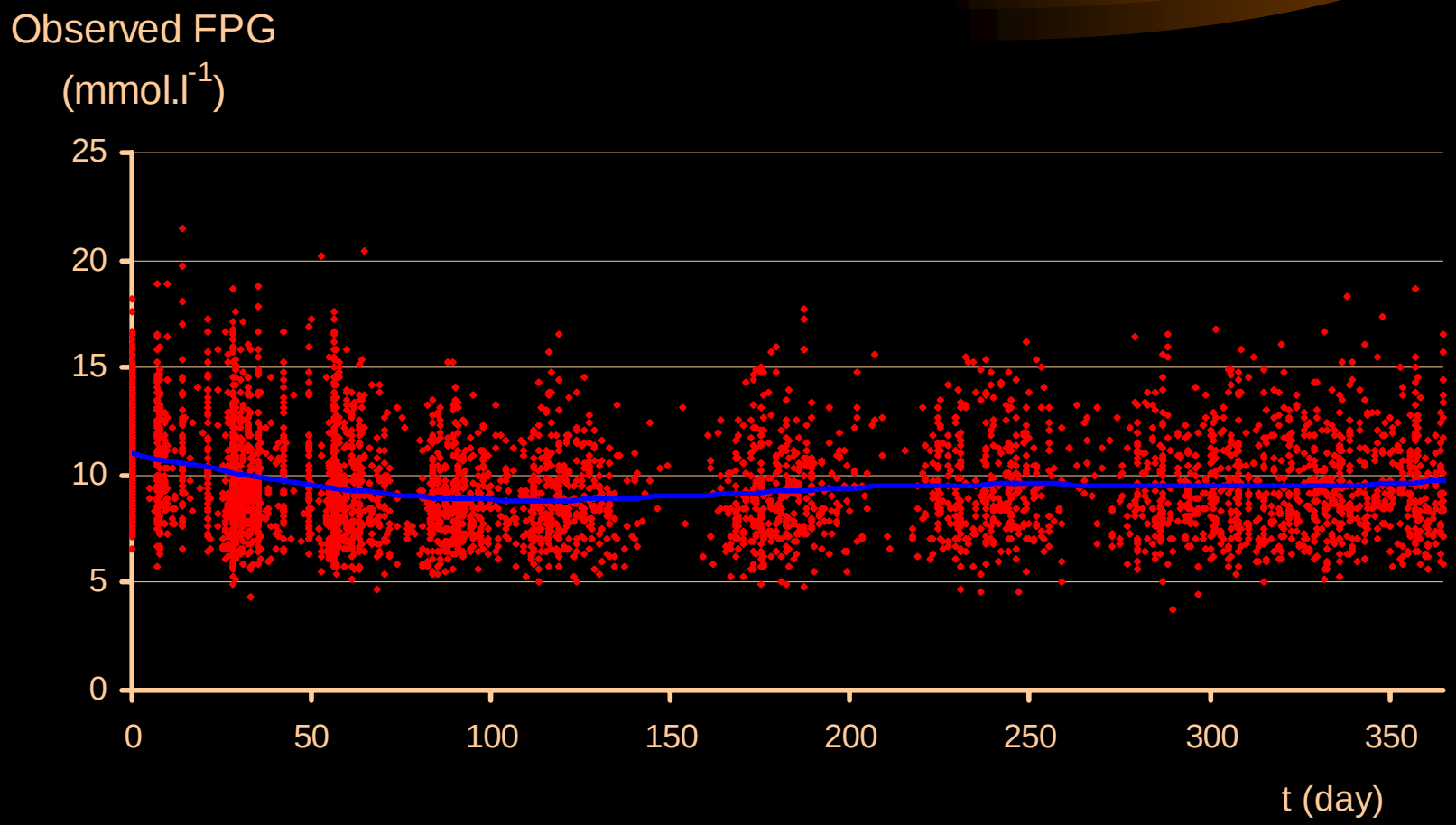
S05702 : Analysis

AUC used to quantify the time course of effect

FPG (fast plasma glucose) used as biomarker

Covariates : age, sex, weight, BMI, CRCL, NBOAD, disdur,…

S05702 : Observed FPG



S05702 : PK/PD model

$$FPG(t)=Base-E(t)+DP$$

with

Base : FPG baseline

DP : Disease progression (function of t)

$$E(t)=P_{max}.Base.AUC_e(t)/ (AUC_e(t)+AUC_{50})$$

AUC_{50}: AUC associated to 50% of effect

$AUC_e(t)$: AUC accumulated in the effect compartment

$$AUC_e(t)=AUC[1-exp(-k_{eq}.t)]$$

k_{eq}: constant of equilibration

S05702 : GH Analysis

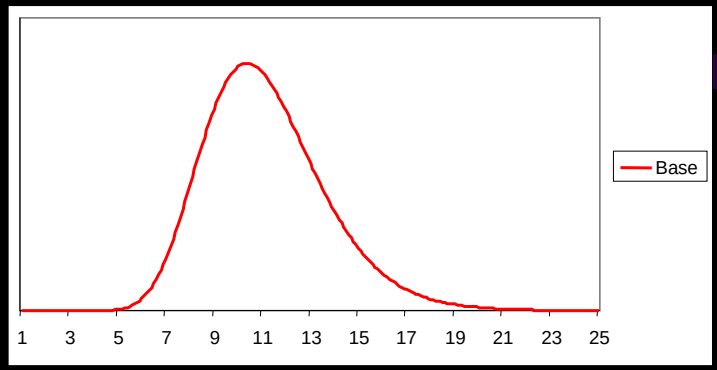
1) Heterogeneity on Base

2) Global heterogeneity

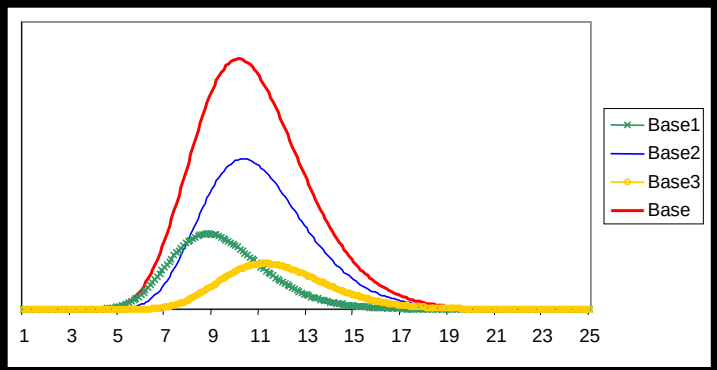
3) Analysis parameter by parameter

4) Final global analysis

S05702 : Results on Base

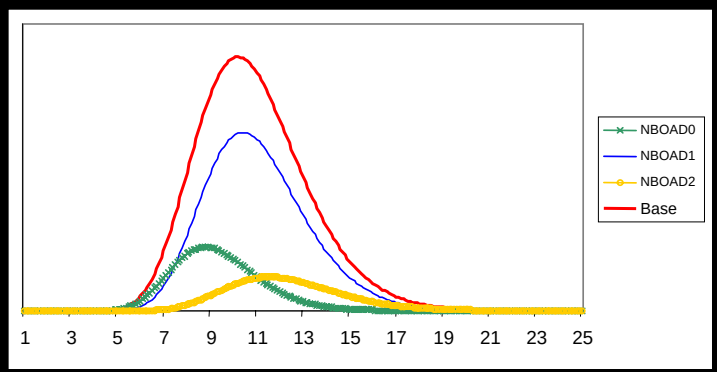


Estimation 1 subpopulation



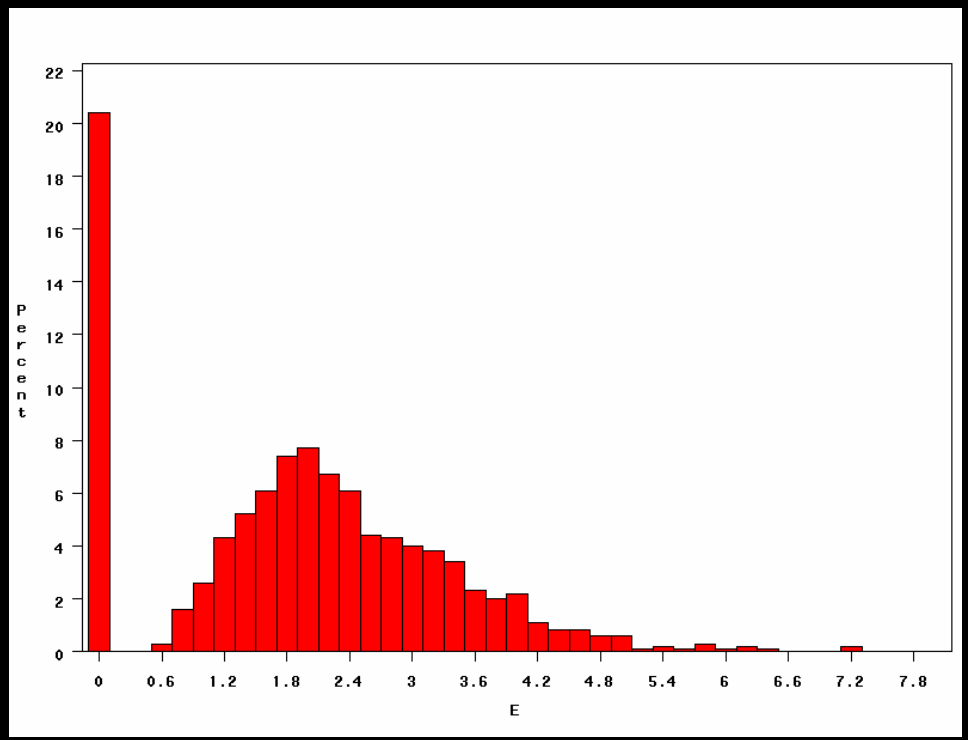
Estimation 3 subpopulations

Logistic Regression -> NBOAD



Estimation 1 subpopulation with the adapted model

S05702 : Results first global analysis



Distribution of the effect E ($mmol.l^{-1}$)

Decrease of FPG

2 subpopulations

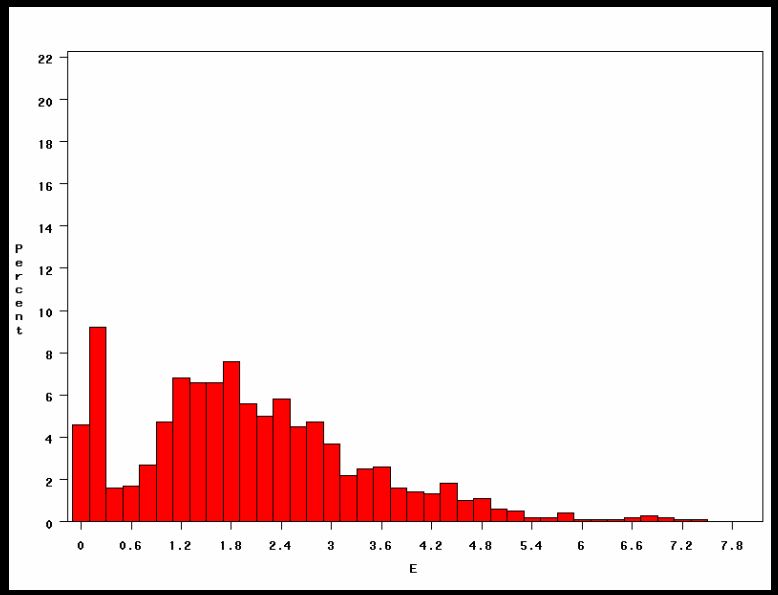
weighted with 21%/79%

One group with very bad responders

S05702 : Final results

1) 2 groups on k_{eq} only partly explained using NBOAD

2) 2 groups on P_{max} explained using NBOAD



3) Final global analysis :

2 subpopulations

weighted with 15%/85%

One group with bad responders

S05702 : Comparison with NONMEM

1) 3 groups of Base ($mmol.l^{-1}$) related to NBOAD

	NBOAD=0	NBOAD=1	NBOAD>1
NONMEM	9.6	10.9	12.2
GH	9.2	10.8	12.9

2) Mixture model with a group of non-responders : E(t)=0, proportion=25%, non-responders related to NBOAD

3) Close estimations of parameters between the 2 methods except for disease progression DP

$DP_{GH}(=1.48\ mmol.l^{-1}) > DP_{NONMEM}\ (=0.84\ mmol.l^{-1})$

Conclusion

Advantages

- Detection of heterogeneities with no assumption of their existence
- Possible explanation of heterogeneities
- Quality of estimation adjustable using the quadrature

Future

- Studying different ways to adapt parameters