

# Evaluation des modèles non-linéaires à effets mixtes

Emmanuelle Comets, Karl Brendel, France Mentré

INSERM UMR738

GDR Statistiques et Santé, 20 octobre 2009

- Définition

- modélisation des données obtenues lors d'essais cliniques sur des médicaments
- développement et application de méthodes mathématiques et statistiques <sup>1</sup> afin de
  - caractériser, comprendre et prédire le comportement d'un médicament
  - quantifier l'incertitude liée à ces éléments
  - permettre des décisions rationnelles

- Discipline en pleine expansion

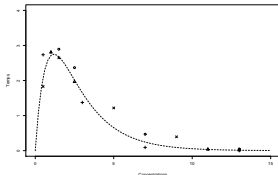
- Modélisation

- modélisation pharmacocinétique (PK)/pharmacodynamique (PD) par des modèles non-linéaires à effets mixtes
- partie intégrante des dossiers d'autorisation de mise sur le marché pour les nouveaux médicaments
- analyse de données répétées recueillies dans les essais cliniques

---

<sup>1</sup> Ette et Williams 2007. *Pharmacometrics: the science of quantitative pharmacology*, Wiley-Interscience.

# Modélisation dans le développement de médicaments



Estimation :

- paramètres  
PK/PD
- variabilité entre  
sujets, ...

Inférence

Optimisation de  
protocoles

Construction de  
modèles  
Estimation (algos)

Evaluation de  
modèles

# Evaluation de modèles (1)

- Etape importante de la construction, pour évaluer les hypothèses faites lors de la modélisation
  - modèle structurel
  - modèle de la variabilité interindividuelle et de la variabilité résiduelle
- Fait l'objet de recommandations
  - recommandation de la FDA concernant les analyses de population dans l'industrie <sup>2</sup>
  - section spécifique sur la façon de rapporter les résultats des études dans la recommandation de l'Agence européenne <sup>3</sup>
- Récente revue bibliographique portant sur 324 modèles PK et/ou PD de population publiés entre 2001 et 2004 <sup>4</sup>

évaluation des modèles jugée bonne ou excellente dans seulement 25% des modèles

---

<sup>2</sup> FDA 1999. *Guidance for Industry: Population Pharmacokinetics*. Rockville, USA

<sup>3</sup> EMEA 2006. *Guideline on reporting the results of population pharmacokinetic analyses*

<sup>4</sup> Brendel et al. *Are population PK and/or PD models adequately evaluated? A 2002 to 2004 literature survey*. *Clin Pharmacokinetics* 2007

# Evaluation de modèles (2)

## ***"No statistical model can safely be assumed adequate" (G. Box, 1980)***

- Différents termes utilisés :
  - validation
  - évaluation
  - adéquation
  - qualification
- Caractéristiques d'un modèle évalué
  - bon ajustement aux données (insuffisant)
  - simplicité (parcimonie)
  - domaine d'applicabilité (extrapolabilité)
- Cependant, pas de consensus sur la meilleure méthode pour évaluer un modèle

# Problématique

Evaluation des  
modèles  
non-linéaires à  
effets mixtes

Emmanuelle  
Comets, Karl  
Brendel, France  
Mentré

Introduction

Métriques pour  
l'évaluation de  
modèles

Construction  
Graphes diagnostiques  
Tests

Etude des  
métriques

Méthodes  
Propriétés statistiques

Conclusion et  
perspectives

- Construction du modèle  $M_B$ 
  - jeu de données  $B$
- Evaluation du modèle à l'aide d'un jeu de données  $V$ 
  - $V=B$  : évaluation interne
  - $V \neq B$  : évaluation externe
- Objectif : tester l'hypothèse

$H_0$  : le modèle  $M_B$  décrit les données de  $V$

# Modèle statistique (1)

- On dispose de données chez  $N$  sujets
- Pour chaque individu  $i$ :
  - $n_i$  observations  $j=1, \dots, n_i$
  - temps de prélèvements :  $t_{ij}, j=1, \dots, n_i$
  - observations
    - $y_{ij}, j=1, \dots, n_i$
    - notation vectorielle :  $\mathbf{y}_i = \vec{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$
- Pour l'ensemble de la population
  - $N$  individus  $i=1, \dots, N \Rightarrow N$  vecteurs d'observations
$$\{\mathbf{y}_i\}_{i=1, \dots, N}$$
  - $n_{tot}$  nombre d'observations au total :  $n_{tot} = \sum_{i=1}^N n_i$

## Modèle statistique (2)

- Modèle à effets mixtes

$$y_{ij} = f(\theta_i, t_{ij}) + \varepsilon_{ij}$$

- vecteur de paramètres individuels  $\theta_i$
- modèle d'erreur

- $\text{var}(\varepsilon_{ij}) = g(\sigma, f(\theta_i, t_{ij}), \beta)$

- Estimation des paramètres de population  $\Psi$ 
  - paramètres de la distribution  $\mathcal{D}$  des paramètres  $\theta_i$
  - paramètres de la variance résiduelle  $\Sigma$
- Modèle  $M_B$  caractérisé par :
  - hypothèses (modèle structural, modèles de variabilité)
  - paramètres estimés  $\hat{\Psi}$



# Erreurs de prédiction (1)

- Différence entre les prédictions et les observations
  - prédictions avec les paramètres de population
    - RES : résidus (prédictions-observations) correspondant
  - prédictions avec les paramètres individuels
    - IRES : résidus individuels
- Pondération pour prendre en compte
  - les corrélations entre les données chez un même sujet
  - les variances attendues
- Erreurs de prédiction normalisées (**spe**)

$$\mathbf{spe}_i = V(\mathbf{y}_i)^{-1/2} (\mathbf{y}_i - E(\mathbf{y}_i))$$

- approximation au premier ordre :
- $E(\mathbf{y}_i)$  : vecteur des prédictions du modèle
- $V(\mathbf{y}_i)$  : matrice de variance-covariance

## Erreurs de prédiction (2)

- Sous l'hypothèse d'un modèle linéaire :

$$spe_{i,j} \sim \mathcal{N}(0, 1)$$

- $spe$  souvent appelées WRES (notation NONMEM) ou CWRES
  - 2 types de linéarisation possibles
- Problème : le modèle n'est pas linéaire!

# Métriques pour les modèles non-linéaires (1)

- Principe : construire la "vraie" distribution des résidus pour un modèle non linéaire
- Construction des erreurs de prédiction sur la distribution
  - distribution prédictive en  $y_{i,j}$  :

$$p_{i,j}(y_{i,j}) = \int_{-\infty}^{\infty} p(y|\theta_i) p(\theta_i) d\theta_i$$

- $F_{i,j}$  : loi cumulative de la densité de probabilité pour l'observation  $y_{i,j}$  :

$$F_{i,j}(y_{i,j}) = \int_0^{y_{i,j}} p_{i,j}(y_{i,j}) dy$$

- erreurs de prédiction sur la distribution pd définies par :

$$pd_{i,j} = F_{i,j}(y_{i,j})$$

- En pratique
  - pas d'expression analytique pour  $p_{i,j}(y_{i,j})$  dans les modèles non-linéaires à effets mixtes
  - approximation stochastique, obtenue en simulant un grand nombre de  $\theta_i$  sous le modèle

# Construction

Evaluation des modèles non-linéaires à effets mixtes

Emmanuelle Comets, Karl Brendel, France Mentré

Introduction

Métriques pour l'évaluation de modèles

Construction

Graphes diagnostiques

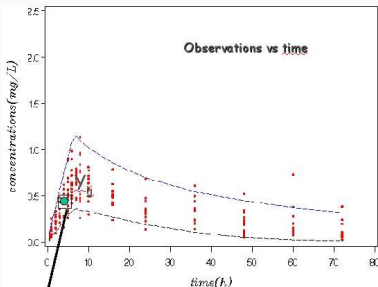
Tests

Etude des métriques

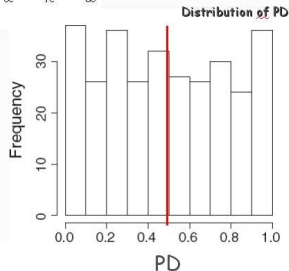
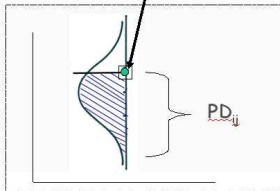
Méthodes

Propriétés statistiques

Conclusion et perspectives



Predicted distribution of observation  $y_{ij}$



# Métriques pour les modèles non-linéaires (2)

- Distribution attendue sous  $H_0$ <sup>5</sup>

$$pd_{i,j} \sim \mathcal{U}(0,1)$$

- Décorrélacion des pd

- calcul de l'espérance  $\mathbf{E}_{emp\ i}$  et de la matrice de variance-covariance empiriques  $\mathbf{V}_{emp\ i}$
- à partir des simulations utilisées pour construire  $p_{i,j}$
- décorrélacion des concentrations observées

$$\mathbf{y}_i^* = \mathbf{V}_{emp\ i}^{-1/2} (\mathbf{y}_i - \mathbf{E}_{emp\ i})$$

- décorrélacion analogue des concentrations simulées :  $\mathbf{y}_{sim\ i}^*$
- $pd_{i,j}^*$  : quantile de  $\mathbf{y}_{ij}^*$  dans la distribution des  $\mathbf{y}_{sim\ ij}^*$
- Normalisation des pd décorrélées : npde
  - soit  $\Phi$  la c.d.f. de  $\mathcal{N}(0,1)$

$$npde_{i,j} = \Phi^{-1}(pd_{i,j}^*) \sim \mathcal{N}(0,1)$$

---

<sup>5</sup>Mentré et Escolano. Prediction discrepancies for the evaluation of nonlinear mixed-effects models. JPKPD 2006

# Propriétés prédictives (PPC, VPC)

- Propriétés prédictives
  - choix d'une statistique calculable à partir des données
  - construction de la distribution de la statistique sous le modèle
  - comparaison à la valeur observée
- Exemple de statistique
  - motifs d'observance dans une étude en données répétées <sup>6</sup>
- Visual Predictive Check (VPC <sup>7</sup>)
  - comparaison entre les données observées et des intervalles de prédiction
  - simulation de jeux d'observation sous le modèle
  - construction de l'enveloppe où devrait se trouver un pourcentage donné d'observations
  - intervalles usuels : 95%, 90%, 50%,...
- Les npde sont des exemples de PPC sur les observations

---

<sup>6</sup> Girard et al. A Markov mixed effect regression model for drug compliance. *Stat Med* 1998

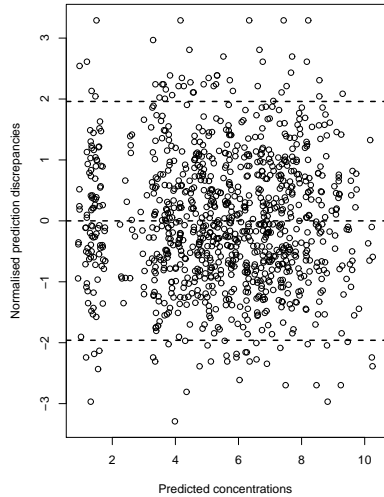
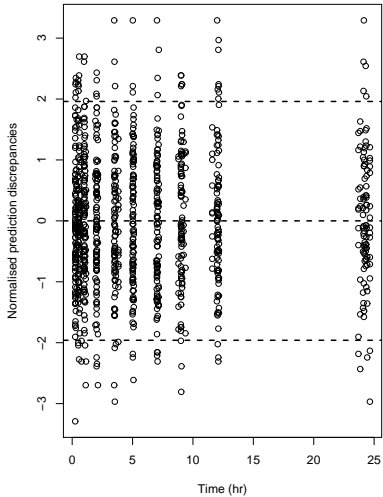
<sup>7</sup> Holford. The Visual Predictive Check: superiority to standard diagnostic (Rorschach) plots. 14th PAGE, Pamplona, Spain, 2005.

# Autres métriques <sup>8</sup>

- Métriques basées sur les hyperparamètres (paramètres de population)
  - erreurs de prédiction standardisées sur les paramètres
  - distribution des paramètres obtenues par simulations
- Métriques liées à la vraisemblance
  - basées sur la différence entre les vraisemblances avec ou sans ajustement
  - distribution obtenue par simulations
- Dans le reste de cette présentation, métriques basées sur des observations continues

---

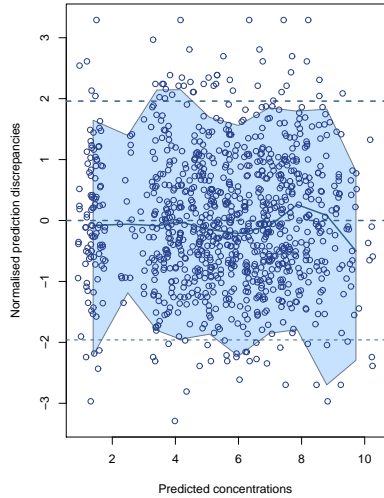
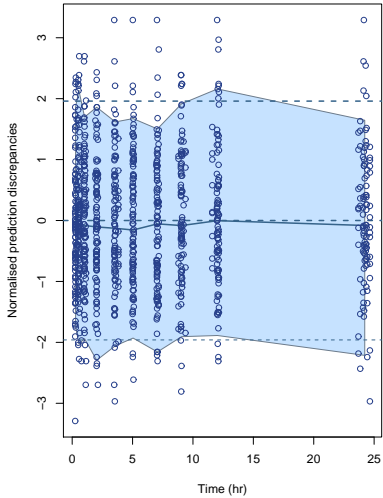
<sup>8</sup>Brendel et al. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. Pharm Res 2006



pd en fonction du temps ou des prédictions



# Graphes des métriques (2)



# Graphes des métriques (3)

Evaluation des modèles  
non-linéaires à effets mixtes

Emmanuelle Comets, Karl Brendel, France Mentré

Introduction

Métriques pour l'évaluation de modèles

Construction

Graphes diagnostiques

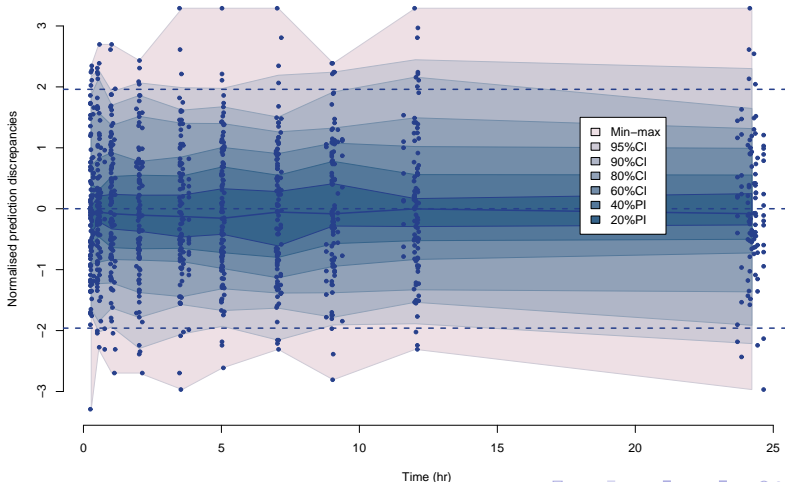
Tests

Etude des métriques

Méthodes

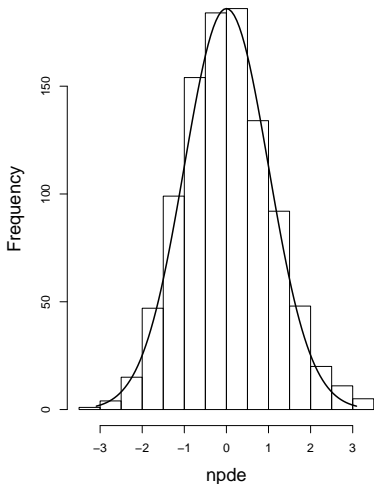
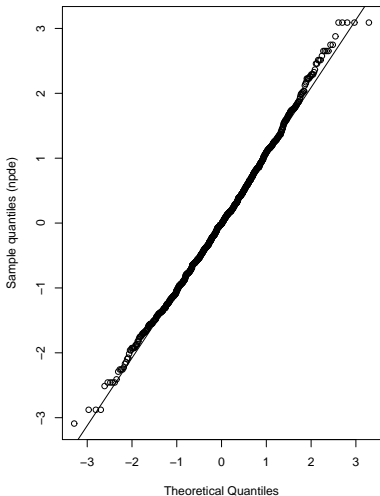
Propriétés statistiques

Conclusion et perspectives



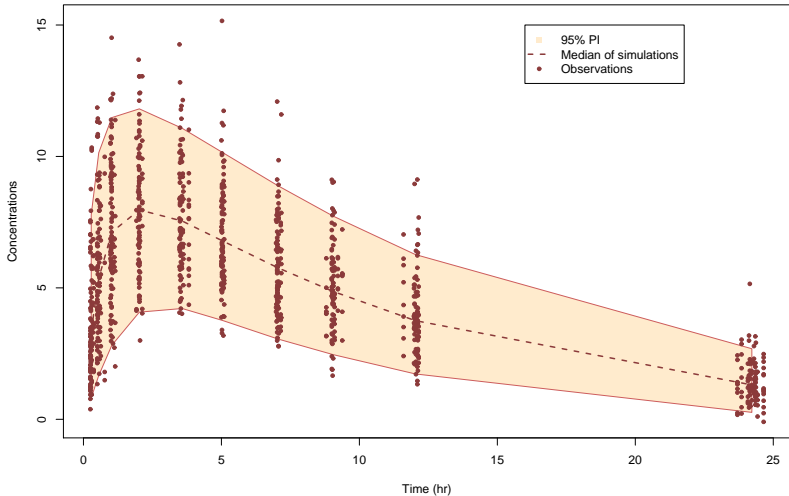
# Distribution des npde

Q-Q plot versus  $N(0,1)$  for npde

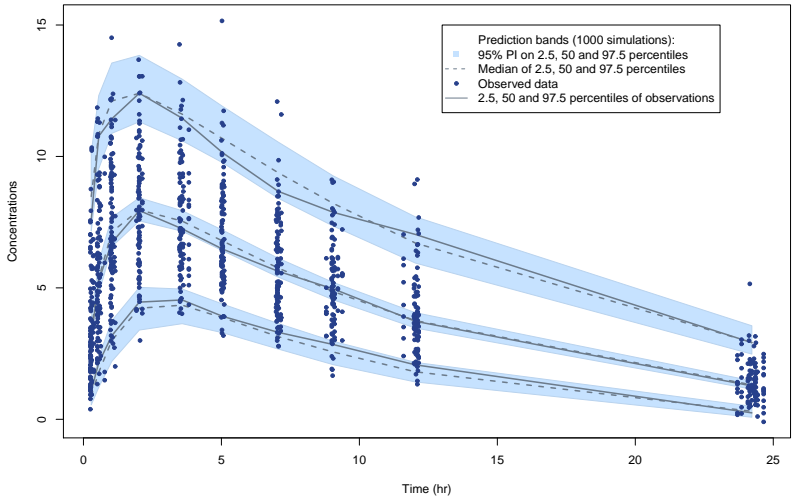


QQ-plot et histogramme de la distribution des npde

# Graphes de VPC (1)



# Graphes de VPC (2)



# Intervalles de prédiction sur pd

Evaluation des modèles non-linéaires à effets mixtes

Emmanuelle Comets, Karl Brendel, France Mentré

Introduction

Métriques pour l'évaluation de modèles

Construction

Graphes diagnostiques

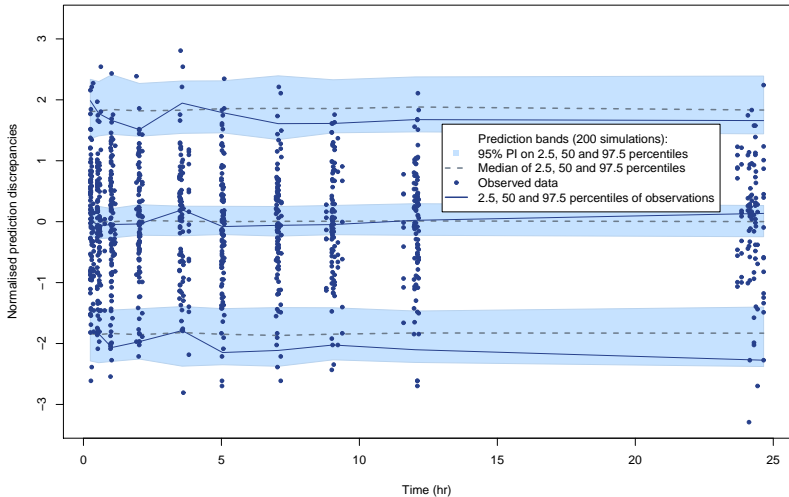
Tests

Etude des métriques

Méthodes

Propriétés statistiques


Conclusion et perspectives



# Tests portant sur la distribution

- Distribution théorique
  - résidus linéarisés normalisés spe
    - suivent  $\mathcal{N}(0, 1)$  sous l'hypothèse de normalité
  - dissonances prédictives pd
    - suivent  $\mathcal{N}(0, 1)$  en l'absence de corrélations
  - erreurs de prédiction sur la distribution npde
    - suivent  $\mathcal{N}(0, 1)$
- Test omnibus
  - test de Kolmogorov-Smirnov
- Tests spécifiques
  - test de normalité
  - test de la moyenne et de la variance
  - test global combinant ces 3 tests avec une correction de Bonferroni<sup>9</sup>

---

<sup>9</sup>Brendel et al. *Evaluation of different tests based on observations for external model evaluation of population analyses JPKPD, sous presse* 

- Visual Predictive Check
  - pas de test formel
  - très utile pour la visualisation des déficiences du modèle
- Numerical Predictive Check (NPC)
  - comparaison du pourcentage de données observées dans un intervalle spécifique au pourcentage attendu sous  $H_0$ <sup>10</sup>
  - intervalles de prédiction usuels : 90%, 80% et 50%
  - test omnibus
    - test de Student
    - test binomial

⇒ choix de l'intervalle ?

⇒ combinaison des résultats (tests multiples) ?
- Classe plus générale: les PPC
  - test de Neymann-Pearson pour toutes les métriques à base de simulations

---

<sup>10</sup> J Wilkins et al. *Patterns and power for the visual predictive check. 15th PAGE meeting, Brugges, Belgium 2006.*



## Exemple d'application

- Gliclazide : médicament antidiabétique
  - données recueillies dans 2 études de phase II et une étude de phase III d'une nouvelle formulation<sup>11</sup>
  - construction du modèle
    - 642 patients souffrant de diabète de type II
    - 5931 concentrations plasmatiques de gliclazide
  - description de la PK par un modèle à un compartiment avec absorption et élimination d'ordre 1
- Jeu de données disponible pour la validation externe
  - deuxième étude de phase III
    - 351 patients
    - 973 observations

---

<sup>11</sup> Frey et al. Population PKPD modelling of the long- term hypoglycaemic effect of gliclazide given as a once-a-day modified release (MR) formulation. *Br J Clin Pharmacol* 2003

# Etude par simulations

Evaluation des  
modèles  
non-linéaires à  
effets mixtes

Emmanuelle  
Comets, Karl  
Brendel, France  
Mentré

Introduction

Métriques pour  
l'évaluation de  
modèles

Construction

Graphes diagnostiques

Tests

Etude des  
métriques

Méthodes

Propriétés statistiques

Conclusion et  
perspectives

- Métriques étudiées
  - résidus linéarisés
  - erreurs de prédiction sur la distribution pd et npde
  - NPC sur les intervalles 90%, 80% et 50%
  - NPC décorrélées sur les mêmes intervalles
    - décorrélation et centrage des données avant le calcul des NPC
- Simulations
  - 500 simulations sous le modèle ( $H_0$ )
  - 1000 simulations pour chaque jeu de données simulé pour le calcul des pd, npde et des VPC
- Evaluation des métriques
  - pour chacune, nombre de jeux de données pour lesquels la métrique/test correspondant rejette le modèle

## Résultats (1)

Test	Métrique		
	spe	pd (norm)	npde
Wilcoxon	4.2%	19.9%	3.0%
Fisher	100%	15.4%	8.6%
SW	100%	16.1%	6.2%
Global	100%	24.6%	5.8%

- Mauvais comportement des métriques linéarisées
  - rejet systématique du modèle selon les spe
- Métriques pd et npde
  - la corrélation entre les données répétées chez un même sujet augmente le risque de rejet du modèle (pd)

## Résultats (2)

Métrique	NPC		NPC décorrélés		
	Test	Student	Binomial	Student	Binomial
Intervalle					
90%		12.0	13.0	7.4	6.2
80%		14.0	13.2	7.8	6.8
50%		13.4	13.4	4.6	4.6

- Les NPC usuels souffrent de la corrélation entre les données
  - décorrélation des NPC par décorrélation des données : revient en fait à calculer les npde
- Tests comparant le pourcentage de données dans l'intervalle par rapport à sa valeur théorique
  - inflation du risque de première espèce avec le test de Student
  - meilleur comportement avec le test binomial
- Problème : multiplicité des tests et/ou choix de l'intervalle

# Discussion

- Métriques pour l'évaluation de modèles
  - métriques linéarisées faciles à calculer
    - mais mauvaises propriétés statistiques
  - nouvelles métriques plus adaptées
    - commencent à être mises en oeuvre dans les logiciels (Monolix)
    - également librairie spéciale `npde` pour les calculer en R <sup>12</sup>:  
[www.biostat.npde.fr](http://www.biostat.npde.fr)
  - VPC très utiles pour évaluer graphiquement un modèle
    - peu informatifs en présence d'hétérogénéité (stratification)
    - pas vraiment de test
  - PPC : évaluation de la capacité prédictive du modèle
    - nécessite de disposer d'une statistique calculable facilement et importante pour l'utilisation des résultats <sup>13</sup>
- Ensemble de graphes et de tests à utiliser
  - évaluation de différents aspects du modèle
  - évaluation adaptée à l'usage qui doit être fait du modèle

---

<sup>12</sup>Comets et al. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Comput Meth Prog Biomed* 2008

<sup>13</sup>Yano et al. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *JPKPD* 2001

- Extensions aux données non continues des métriques basées sur des simulations
  - données binaires ou catégorielles <sup>14</sup>
  - données de comptage
- Génération des simulations possiblement problématique <sup>15</sup>
  - gestion/modélisation de données censurées, perdus de vue
  - protocole adaptatif ou observationnel
- Représentation graphique <sup>16</sup>
- Evaluation *versus* validation
  - modèle non rejeté  $\neq$  modèle accepté
  - notion de modèle acceptable
  - construction d'intervalles de confiance autour de la distribution

---

<sup>14</sup> Hénin et al. A dynamic model of hand-and-foot syndrome in patients receiving capecitabine. *Clin Pharmacol Ther* 2008

<sup>15</sup> Karlsson et Savic. Diagnosing model diagnostics. *Clin Pharmacol Ther* 2007

<sup>16</sup> Laffont et Concordet. A new exact test to globally assess a population PK and/or PD model. 18th PAGE meeting, St-Petersburg, Russia 2009.