

Are population PK and/or PD models adequately evaluated?

A survey of the literature between 2002 to 2004

K. Brendel⁽¹⁾, C. Dartois⁽²⁾, E. Comets⁽¹⁾, A. Lemenuel-Diot⁽³⁾, C. Laveille⁽⁴⁾, B. Tranchand^(2,5), P. Girard^(2,6), C. Laffont⁽³⁾ and F. Mentré⁽¹⁾.

(1) INSERM U738, Paris, France; (2) EA3738, Lyon, France; (3) IRIS Servier, Courbevoie, France; (4) Exprimo, Lummen, Belgium; (5) CLB, Lyon, France; (6) INSERM, Lyon, France

Introduction

- Evaluation for non linear mixed-effects models is a complex issue in statistical modeling and is called by different terms:
 - Evaluation, validation, qualification, adequacy, assessment, checking, appropriateness, performance...
- There is no consensus on a general approach to evaluate a population model for **PK** and/or **PD** analyses
- FDA has stressed the need for model evaluation (1999)

Objectives of the review

- Primary objective
 - Survey the different methods currently used to evaluate **PK** and/or **PD** population models
- Secondary objective
 - Assess whether those models were adequately evaluated

Methods

- Articles selection
- Data abstraction form (DAF) building
- Data collection
- Population model evaluation section in the DAF

Methods

- Population model evaluation section in the DAF

Evaluation methods divided into 3 subsections in the DAF

- Basic internal evaluation
(Goodness-of-fit plots (GOF), uncertainty on parameter estimates)
- Advanced internal evaluation
(Data splitting, resampling techniques, Monte Carlo simulations)
- External evaluation (predictability of the model on the external dataset)

Subjective synthesis

1. Was there an attempt to evaluate the model?
2. Was the type of evaluation justified?
3. Was the choice of the metrics appropriate?
4. Was the model evaluated?

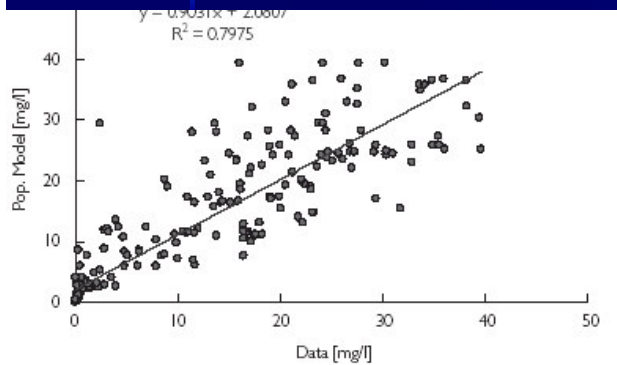
4 possible answers: no, poor, good, excellent

Results: population model evaluation (360 PK models, 118 PD models)

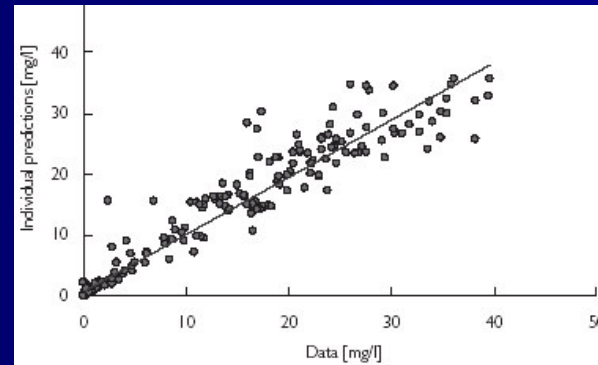
Type of evaluation	PK	PD
Basic internal	74%	76%
Advanced internal	28%	16%
External	7%	8%

Basic internal evaluation

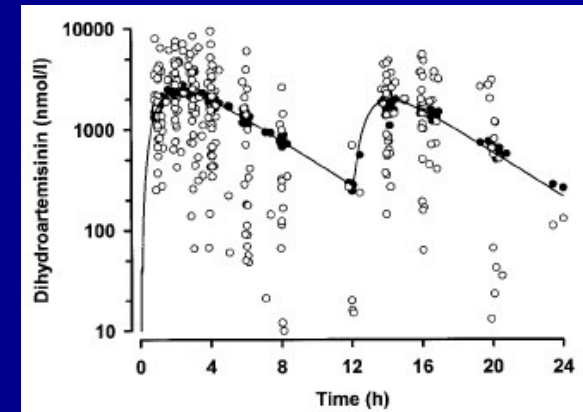
■ GOF used and shown (65%, 60%)



PRED vs OBS
(36%, 25%)



IPRED vs OBS
(29%, 14%)



OBS vs time
with mean PRED
(26%, 35%)

Basic internal evaluation

- Uncertainty on parameter estimates (SE or CI):

	<i>Estimate</i>	<i>Standard error</i>
β_1	0.023	0.0016
β_2	0.000 057	0.000 055
β_3	0.57	0.053
Interpatient variability in clearance	0.15 (38.7%)	0.038

For fixed effects
(53%, 42%)

For random effects
(26%, 11%)

- Uncertainty mostly obtained by:
 - Fisher information matrix (78%, 87%)

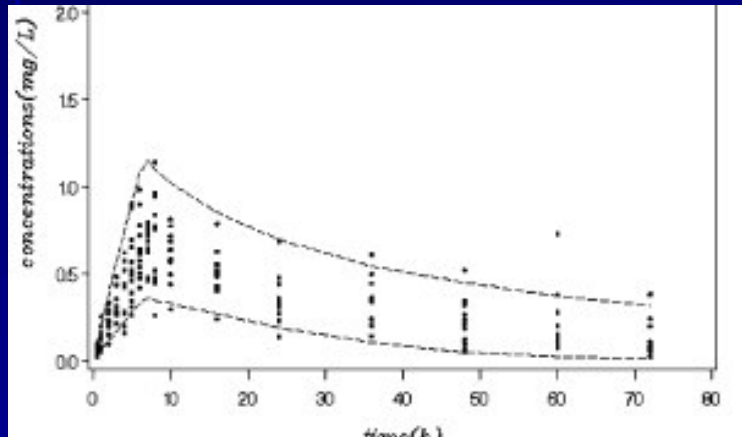
Advanced internal evaluation

- Data splitting (13%, 3%)
 - With randomization (7%, 0%)
 - Size of validation dataset/entire dataset, median (1/3, 1/4)
- Bootstrap (9%, 5%)
 - Number of bootstrap samples, median (500, 200)
- Cross validation (3%, 4%)
 - Number of groups, median (10, 13)

Advanced internal evaluation

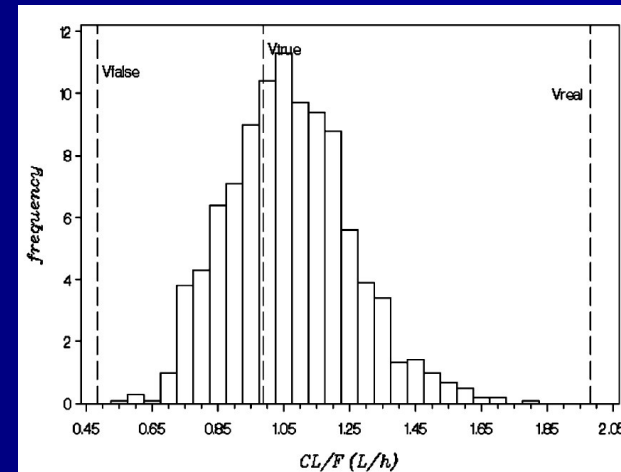
- Monte Carlo simulations (8%, 6%)

- Aim of simulations:



Visual Predictive Check

(3%, 3%)



Statistical Predictive Check

(2%, 0%)

External evaluation (7%, 8%)

- External dataset obtained from studies with less than 50 patients
- Same inclusion criteria (3%, 6%)
- Same dose regimen (2%, 0%)
- Same sampling scheme (1%, 5%)

Metrics used for model evaluation

- For observations:
 - Prediction errors (PE or RES) and standardized prediction errors (SPE or WRES) were the most used metrics
 - From these metrics, bias (mean of PE or MPE) and root mean square errors (RMSE) were the most reported
- For individual and population parameters:
 - MPE was the most reported

Subjective synthesis

1. Was there an attempt to evaluate the model?

no/poor (62%, 68%); good/excellent (38%, 32%)

2. Was the type of evaluation justified?

no/poor (79%, 86%); good/excellent (21%, 14%)

3. Was the choice of the metrics appropriate?

no/poor (72%, 74%); good/excellent (28%, 26%)

4. Was the model evaluated?

no/poor (72%, 74%); good/excellent (28%, 26%)

Conclusion

- The first large review of literature for population model evaluation
 - SE or CI reported in only 1/2 of the models for fixed effects and even less for random effects
 - Advanced methods used only in 1/4 of analyses
 - Data splitting was the technique the most used among the advanced methods
- Model evaluation was judged good or excellent in only 1/4 of the models

Acknowledgments

- All the authors who have published papers between 2002 and 2004...