Tests based on objective functions for external evaluation of population pharmacokinetic models: illustration on the population PK of gliclazide

K. Brendel[1], E. Comets[1], C. Laffont[2], C. Laveille[2], F. Mentré[1]

[1]INSERM U738, Department of Epidemiology, Biostatistics and Clinical research, AP-HP, Bichat University Hospital, Paris, France,[2]Servier,Courbevoie, France

INTRODUCTION

An important step in population pharmacokinetic model building is to evaluate the model's adequacy. Two types of evaluation can be performed. The first is internal evaluation and refers to the use of data splitting and resampling techniques; in the following, we only consider the second, external evaluation which refers to a comparison between the validation dataset and the predictions from the model built using the learning dataset (B). The validation dataset is not used for model building and parameters estimation. Several types of prediction errors on concentrations, random effects and hyperparameters were proposed and evaluated on simulated datasets [1]. We propose here three tests based on objective functions for external evaluation. We constructed a population model of gliclazide and we then applied the tests to four validation datasets: three simulated datasets and one dataset from a phase I study.

METHODS

Let M^B be the model and population parameters to be evaluated and V a validation dataset. The model was built from 2 phase II studies of gliclazide, an antidiabetic drug. M^B was a one compartment model with zero order absorption and first order elimination, with exponential random effects on the apparent volume of distribution (V/F) and on the apparent clearance (CL/F). A proportional error model was selected.

Hyperparameter	Estimate	SE
CL/F (L/h)	1.0	0.042
V/F (L)	39.8	2.3
Tabs (h)	6.6	0.22
$\omega^2_{CL/F}$	0.35	0.057
$\omega^2_{V/F}$	0.11	0.028
σ^2	0.06	0.0064

Table 1 Estimated population pharmacokinetic parameters of gliclazide with M^B.

Three validation datasets were simulated according to the design of a real phase I study with 12 subjects and 16 samples (V_{real}): the first (V_{true}) was simulated using the parameters of M^B; the second and third datasets were simulated using the same model, but with a bioavailability multiplied by two (V_{false1}) or divided by two (V_{false2}).

Values below the quantification limit (BQL) were treated by imputing the last BQL measurement to BQL/2 and omitting previous BQL values during the ascending phase and inversely in the descending phase.

We defined three tests applied on metrics based on objective function (OF) for model evaluation. We considered metrics without and with simulations. These last metrics, called posterior predictive check (PPC), evaluate the adequacy between data and model by comparing a given statistic, computed with the data, to its posterior predictive distribution computed under the model. This distribution was estimated using Monte Carlo simulations with M^B to obtain K datasets simulated according to the phase I design.

We simulated these three datasets to check the ability of the metrics to validate V_{true} and to reject V_{false1} and V_{false2}. An illustration of the metrics is finally done on V_{real}.

Metrics based on objective functions

OF can be determined with two methods: first, with model M^B and hyperparameter Ψ^B without fitting the dataset V (OF_{nofit}^V; all parameters fixed), second with hyperparameter Ψ^V after fitting the model on the dataset V (OF_{fit}^V; all parameters estimated). Several metrics can be defined from these objective functions, without and with simulation.

Prediction Error on Gain in Objective Function (PEGOF)

We can define the prediction error as:

$$PEOF = \Delta OF^V = OF_{nofit}^V - OF_{fit}^V$$

Prediction Error on Objective Function with Simulation (PEOFS)

OF_{nofit}^V can also be compared to the posterior predictive distribution of the objective function estimated from K Monte Carlo simulated datasets with M^B, yielding a value OF_{nofit}^{simk}.

Prediction Error on Gain in Objective Function with Simulation (PEGOFS)

A third approach compares the ΔOF^V with its posterior predictive distribution. For each k simulated dataset, we calculate the gain in objective function with M^B ($\Delta OF^{simk} = OF_{nofit}^{simk} - OF_{fit}^{simk}$) which is then compared to ΔOF^V.

As the simulated datasets may have different number of values below the limit of quantification (BQL), they may have different number of observations after treating the BQL.

Tests and graphs

The difference between two hierarchical models is approximately chi-square distributed. For PEGOF, we compare a model with 0 parameters (all parameters fixed) with a model with Q parameters (all parameters estimated). So we can test model adequacy by comparing the PEGOF to the critical value of a chi-square with Q degrees of freedom.

For tests applied to PEOFS, the K values of OF_{nofit}^{simk} are sorted and the percentile of OF_{nofit}^{simk}, perc, is defined as the number OF_{nofit}^{simk} of below OF_{nofit}^V divided by K. The p-value of the two sided test based on the empirical distribution can be calculated as:

$$p = 2 \times \min(perc, (1 - perc))$$

For PEGOFS, to compare ΔOF^V with the empirical distribution of ΔOF^{simk}, as OF_{nofit}^V is necessarily higher or equal to OF_{fit}^V we have to calculate the p-value of an unilateral test as:

$$p = (1 - perc)$$

The p-value is then compared with 0.05.

To illustrate PEOFS and PEGOFS, histograms of the predictive distribution of OF_{nofit}^{simk} or ΔOF^{simk} can be done and the estimated value on V, OF_{nofit}^V or ΔOF^V are added.

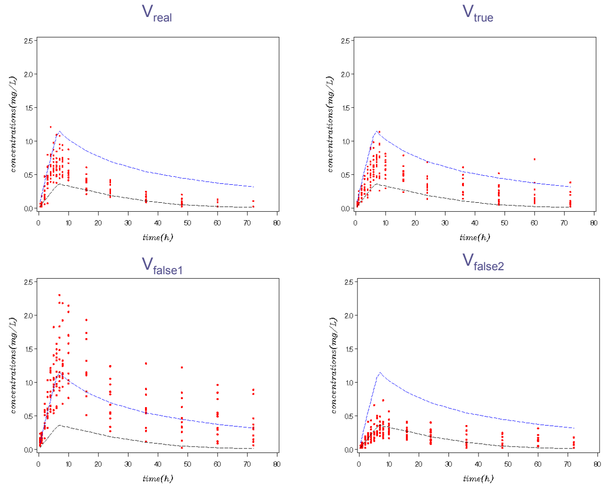

Figure 1 The four validation datasets: the dashed lines represent the 80 % predicted interval, obtained for each time-point as the 10th and 90th percentiles of 1000 simulations under M^B.

RESULTS

The four validation datasets are displayed in Figure 1.

For V_{true}, the objective function with M^B is –751 without fitting and –754 with fitting. The gain in OF from fitting (3) is not significant (p=0.86) according to a LRT with Q=6 degrees of freedom. Compared to the predictive distribution in the simulated datasets we do not reject V_{true} both for OF_{nofit}^V (p=0.092) and for ΔOF^V (p=0.90).

For V_{false1}, the objective function with M^B is –421 without fitting and –474 with fitting. The gain in OF from fitting (53) is significant (p<0.0001) according to a LRT with Q=6 degrees of freedom. V_{false1} is rejected for OF_{nofit}^V (p<0.0001) compared to its predictive distribution in the simulated dataset and is also rejected for ΔOF^V (p<0.0001).

For V_{false2}, the objective function with M^B is –807 without fitting and –859 with fitting. The gain in OF from fitting (52) is significant (p<0.0001). Compared to the predictive distribution in the simulated datasets we reject V_{false2} both for OF_{nofit}^V (p<0.0001) and for ΔOF^V (p<0.0001).

Histograms of the predictive distribution of the objective function without fitting and for are displayed in Figure 2. The three metrics performed similarly on the validation datasets simulated by not rejecting V_{true} and by rejecting V_{false1} and V_{false2}.

As an illustration, for V_{real}, the objective function with M^B is –600 without fitting and –661 with fitting. The qain in OF from fitting (61) is significant (p<0.0001) according to a LRT. V_{real} is not rejected for OF_{nofit}^V (p=0.062) compared to its predictive distribution in the simulated dataset but is rejected for ΔOF^V (p<0.0001).

The empirical posterior distribution for PEOFS does not correct for the varying number of data involved in each simulated dataset and we think it is then preferable to compare the observed gain of objective function on the simulated dataset.

Datasets	PEGOF p-value	PEOFS p-value	PEGOFS p-value
V_{true}	0.86	0.092	0.90
V_{false1}	<0.0001	<0.0001	<0.0001
V_{false2}	<0.0001	<0.0001	<0.0001
V_{real}	<0.0001	0.062	<0.0001

Table 2 P-values of the 3 metrics based on objective function (PEGOF, PEOFS and PEGOFS) obtained with M^B applied to the four validation datasets.

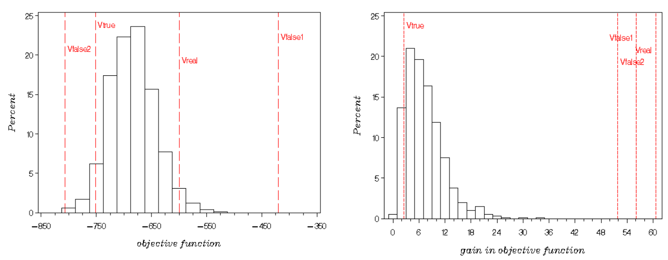

Figure 2 Histogram of the objective functions when M^B is applied to 1000 datasets without estimation (left panel). The values of the corresponding values found for V_{true}, V_{false1}, V_{false2} and V_{real} using M^B are shown as dotted lines. Histogram of the gain in objective functions when M^B is applied to 1000 datasets without and with estimation (right panel). The values of the corresponding values found for V_{true}, V_{false1}, V_{false2} and V_{real} using M^B are shown as dotted lines.

CONCLUSION

The metrics based on objective function are an interesting tool for external evaluation. The method without simulation is very simple and, in this example efficient to detect model inadequacy. Metrics based on the gain in objective function, PEGOF and PEGOFS, showed model misfit and are more adapted than PEOFS by taking into account the number of values below the limit of quantification. For metrics with simulation, we found the same results, but simulation is much more cumbersome because all simulated datasets have to be fitted. A next step will be to evaluate the performance of these metrics with extensive simulation.

Reference: [1] K. Brendel, E. Comets, C. Laveille, R. Jochemsen, F. Mentré. Comparison of several prediction errors for model evaluation. 13th Meeting of the Population Approach Group in Europe, Uppsala (2004)