CAN NON PARAMETRIC METHODS IMPROVE SUB-POPULATIONS DETECTION ?

A simulation-based comparison of Non Parametric (NP) methods for population PK analysis

J. Antic[1,2], C.M. Laffont[1], D. Chafaï[2], D. Concordet[2]

[1]Servier Research Group, Clinical Pharmacokinetics Department, Courbevoie, France ,[2]INRA/ENVT Joint Research Unit 181, Physiopathology and Experimental Toxicology, Toulouse, France

INTRODUCTION

Presence of sub-populations not explained by available covariates casts doubts on the assumption of unimodality made by classical population parametric methods (FO, FOCEI), and on the ability of these methods to identify such sub-populations.

Non Parametric (NP) methods make no assumption on the population distribution, which makes them attractive. However, little is known on the practical performance of these methods. The aim of the present study is thus to evaluate the usefulness and feasibility of NP methods for population approaches, based on simulations with a simple PK model.

Is the EBE's empirical distribution a reliable tool for subpopulation detection ?

Classical population estimation methods (FO, FOCEI…) deals with (log-)normality and unimodality assumptions. The assumption of unimodality is usually checked *a posteriori* by inspection of Empirical Bayes Estimates (EBEs). However, EBEs highly depend on the normality assumption, particularly with sparse datasets. Then EBEs histogram as a normality diagnostic is more and more questioned [1].

What are the cost and benefits of NP methods compared to their counterpart (EBEs)?

Being assumption-free, NP methods seem better suited for identifying sub-populations. However, NP methods are known to be computational expensive, while their practical performances are not well documented.

Are NP methods (NPML, NPEM, NP-NONMEM) equivalent ?

In the 80's, NonParametric Maximum Likelihood (NPML) [2] and NonParametric Expectation Maximisation (NPEM) [3] algorithms were proposed to compute the NP Maximum Likelihood Estimator (NP-MLE). More recently, NONMEM VI introduced a nonparametric step based on EBEs (NP-NONMEM). To the best of our knowledge, no comparison between these methods has been done.

SEVERAL NON PARAMETRIC METHODS

◆ DISTRIBUTION OF EMPIRICAL BAYES ESTIMATES (EBEs)

Empirical distribution of the EBEs' following parametric estimation (ex: FOCEI) can be seen as a discrete nonparametric estimation of the population distribution (with EBEs as support points and equal frequencies), which leads to a **consistent estimation provided the number of observations per individual is large.**

◆ NONPARAMETRIC NONMEM VI (NP-NONMEM)

Like the EBEs histogram, NP-NONMEM estimator is a discrete measure with EBEs used as support points. Difference between EBEs histogram and NP-NONMEM rests in probabilities assigned to each support point : they are all equal for the EBEs empirical distribution, whereas NP-NONMEM computes them by likelihood maximization [4].

NP-NONMEM theoretical properties are, as yet, unknown.

◆ NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATOR (NP-MLE)

When no constraint is imposed on the distribution, for most population models, NP-MLE is a **consistent estimation even with sparse data** [5].

Besides, like the above NP estimates, NP-MLE is a discrete distribution with at most as many support points as individuals in the sample [6]. However, support points can be different from EBEs. Support points and the associated probabilities should be computed by likelihood maximization. Unfortunately likelihood optimization is too difficult to have an explicit solution.

NPML [2] and NPEM [3] are algorithms to compute the NP-MLE. Given an initial distribution, they improve the likelihood at each iteration. So, provided that the initial distribution is close enough from the NP-MLE, NPML and NPEM reach the NP-MLE.

SIMULATIONS LAYOUT

◆ MODEL

A monocompartimental IV *bolus* model with a proportional residual error model (20%).

Distribution was log-normal for V and a mixture of 2 log-normal distributions for CL (see figure 1).

Population means were 1.46 L for volume of distribution (V) and 7.1 mL/min for clearance (CL). Between-Subject Variability (BSV) was 16% for V and 19% for CL (intra sub-population BSVs were 15.8% and 6.30%). CL and V were independent.

◆ EXPERIMENTAL DESIGNS

- One single drug administration (25 mg)
- Different sample sizes : 50, 100 or 300 individuals.
- For 20 individuals of a sample (whatever the sample size), 6 observations/individual were available from 1 min to 5 h after drug administration
- For the others individuals of the sample, only 1 observation/individual was available. These observations were half early, half late (respectively around 2 min and 4 h after administration).

For each sample size, 100 different datasets were simulated and analyzed using NONMEM VI FOCEI and the NP methods (NP-NONMEM, NPML, NPEM). EBEs empirical distribution was chosen to initialize all NP methods.

RESULTS

- **EBE's histogram generally failed to detect the two subpopulations for CL. In contrast, NP methods often seem to detect the bimodality.** These results are presented in Figure 1.

- However, NP methods frequently seem to evidence multimodality for V, while V was simulated according to a log-normal distribution (Figure 2).

- Computation time was a few seconds for NONMEM methods, few minutes for NPML and nearly one hour for NPEM, for samples of 300 individuals (less for smaller samples).

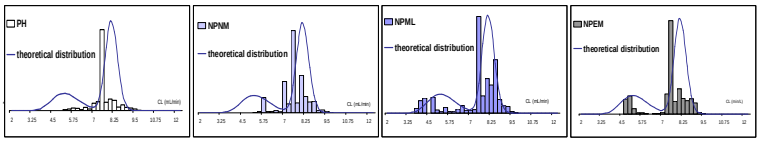

Figure 1 : Theoretical vs estimated distribution of clearance with : EBEs empirical distribution, NP-NONMEM, NPML and NPEM, for 300 individuals sample.

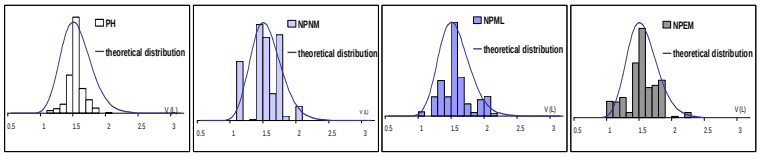

Figure 2 : Theoretical vs estimated distribution of volume of distribution with : EBE's empirical distribution, NP NONMEM, NPML and NPEM, for a 300 individuals sample.

◆ PERFORMANCE CRITERION

How to measure similarities between estimated and true distributions ?

Many distances are available : quantiles, Kolmogorov-Smirnov, T1…

T1 distance seems interesting for this study because it is defined as the integral of the absolute difference between the cumulative distribution functions (cdf) and is equivalent to the average of the absolute difference between quantiles functions.

T1 distances were computed between the true and estimated marginal cdf. Boxplots of these T1 distances for all samples were drawn with respect to the sample size in order to illustrate performances of the methods: for a consistent estimation, the distance should decrease to 0.

- EBEs: as T1 distance increases with the sample size, EBEs distribution appears as a non consistent estimation, in particular in the case of CL (bimodal distribution).
- NP estimates: T1 distance decreases with the sample size. Thus, these estimates could be consistent. Further investigations (larger samples) would be necessary to clarify their properties. Based on our simulations, differences between NP-NONMEM and NP-MLE are small with respect to T1 distance.

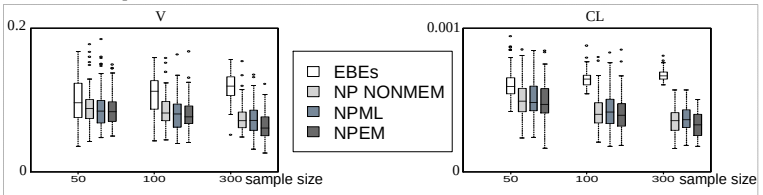

Figure 3 : Boxplots of T1 distances between true and estimated marginal cdf with respect to the sample size

DISCUSSION & CONCLUSIONS

In our simulations, **EBEs histogram was a poor tool for subpopulations detection.**

In contrast, **NP methods were more inclined in sub-populations identification.** However, as they often tended to exhibit multimodality even when real distribution was unimodal, **NP methods could lead to over-doubt unimodality.**

For NP-MLE methods (NPML, NPEM), computational cost was real. Actually, **computation of the NP-MLE remains a very tricky task.** Computation time cost and numerical traps could put NPML and NPEM at a real disadvantage with more complex population models or larger samples. A new NP-MLE algorithm [7], especially designed for high dimension, would deserve some investigation.

On the contrary, **NP-NONMEM's computation time was hardly longer than FOCEI's**. Besides, regarding the T1 distance, even if NPEM seems slightly better than NP-NONMEM for large sample sizes, **NP-NONMEM practical performances were quite good. However, since its statistical properties are not known, such a good behavior in other models is not guarantee...**

References

[1] Savic RM, Wilkins JJ, Karlsson MO. (Un)informativeness of Empirical Bayes Estimate-Based Diagnostics. In: The AAPS Journal; 2006; Vol. 8, No. S2, Abstract T3360 (2006): American Association of Pharmaceutical Scientists; 2006.**[2]** Mallet A., A maximum likelihood estimation method for random coefficient regression models. Biometrika, 73(3) : 645-656,1986. **[3]** Schumitzky A., Nonparametric EM algorithms for estimating prior distributions. Applied Mathematics and Computation, 45(2, partII) : 143-157, 1991. **[4]** Savic R., Kjellson M., Karslsson M., Evaluation of the nonparametric estimation method in NONMEM VI beta. PAGE meeting, 2006. **[5]** Chafaï D., Loubes J.M. On nonparametric maximum likelihood for a class of stochastic inverse problems.Statistics & probability letters, vol 76, issue 12,1225-1237, 2006 **[6]** Lindsay B. G., The geometry of mixture likelihoods: a general theory. The annals of statistics, 11(1):86-91, 1983 **[7]** Leary R., An evolutionary nonparametric NLME algorithm. PAGE meeting 2007.

✉ j.antic@envt.fr