

Title: Normalised prediction distribution errors for the evaluation of nonlinear mixed-effects models: development of the npde add-on package for R

Authors: Emmanuelle Comets (1) , Karl Brendel (2) and France Mentré (1,3) *

Institutions: (1) INSERM, U738, Paris, France; Université Paris 7, UFR de Médecine, Paris, France; (2) Institut de Recherches Internationales Servier, Courbevoie, France. (3) AP-HP, Hôpital Bichat, UF de Biostatistiques, Paris, France

Objectives: Model evaluation is an important part of model building. Mixed-effect models are increasingly used in the analysis of longitudinal data from clinical trials, especially in the area of population pharmacokinetics (PK)/ pharmacodynamics (PD). Standardised prediction errors have long been used to diagnose model misspecifications, but they involve a linearisation of the model which is not appropriate for nonlinear models. Prediction discrepancies (pd), have been proposed by Mentré and Escolano [1] to evaluate nonlinear mixed effect models. They rely on the predictive distribution for each observation, approached through extensive Monte-Carlo simulations, and thus do not involve approximations. In a previous work [2], we developed an improved version of this metric, termed normalised prediction distribution errors (npde), taking into account repeated observations within one subject. The simulations involved in the computation of npde are the same as those needed to perform visual predictive checks (VPC) [3] which are being increasingly used as a model assessment tool and relate to the more general category of posterior predictive check (PPC).

In the present paper, we present a set of routines, written as an add-on package for the open-source statistical package R[4], to compute npde. To illustrate the use of npde, we consider the well-known theophylline PK dataset and simulate two datasets based on the same design that we use as external validation data.

Methods: Model evaluation consists in assessing whether a given model M (composed of a structural model and population parameter estimates) adequately predicts a validation dataset V . V can be the original dataset used to build model M (internal validation) or a separate dataset (external validation). The null hypothesis H_0 is that the data in V can be described by model M . The pd for a given observation is defined as the percentile of this observation within the marginal predictive distribution for that observation, under the null hypothesis (H_0) that the model under scrutiny adequately describes a validation dataset. The predictive distribution is obtained assuming the posterior distribution of the estimated parameters by maximum likelihood estimation, disregarding the estimation error (plug-in approach). Prediction distribution errors are computed in a similar way after correcting for the correlation induced by repeated observations within one individual. The predictive distribution itself is approximated by Monte-Carlo simulations: K datasets are simulated under the null hypothesis (model M and corresponding parameters) using the design of V . Under H_0 , the npde follow the standard normal distribution $N(0,1)$. We test this hypothesis using a combination of 3 tests: a Wilcoxon signed rank test for the mean, a Fisher variance test, and a Shapiro-Wilks normality test. We also propose a global test, involving the Bonferroni correction on the 3 previous tests.

We applied the library to the well-known theophylline PK dataset, including 12 patients given a single dose of theophylline and sampled 10 times each over 12 hours. We illustrate the use of the package with two simulated datasets, one under the true model (V_{true}) and one with the same model but a mean volume of distribution multiplied by 2 (V_{false}).

Finally, we evaluate the influence of K on the results of the tests, for various number of patients. To simplify computations we assume the same sampling design for all subjects, and we simulate datasets

with 12, 100, 250 and 500 subjects with the two models used to simulate V_{true} and V_{false} . We compute the npde for each dataset with K varying from 100 to 1000.

Results: The add-on package npde, developed in R, requires as input a file with the validation dataset V and a file containing the K simulated datasets stacked one after the other. Simulations should be performed beforehand which can be done for instance with the software used for estimation of the population parameters. The program then computes the npde. Optionally, pd can be computed instead or in addition, which is less time-consuming but leads to type-I error inflation especially as the number of observations per subject increases. Graphical diagnostics are plotted to evaluate model adequacy. Tests are performed to compare the distribution of the npde relative to the expected standard normal distribution. The code is available on the package's website <http://www.npde.biostat.fr/>

Applied to the datasets V_{true} and V_{false} , simulated with the same design as the theophylline dataset, the npde reject V_{false} but don't show any problem with V_{true} , as expected. We show the QQ-plots of the npde versus the normal distribution, a histogram of the npde with the density function of the theoretical distribution overlaid, and we also provide plots of the npde versus time or predicted concentrations, that can be used to help model diagnosis.

Finally, the simulations with varying K show that K needs to be larger for large datasets, and we suggest to use at least $K=1000$ for medium-sized datasets, while large datasets appear to need $K=5000$.

Conclusions: The npde can be seen as a numerical predictive check, which can complement more graphical diagnostics such as VPC. There is no clear test for VPC, and in addition multiple observations per subject induce correlations in the VPC. On the other hand, npde are decorrelated, and should follow the expected standard normal distribution if the model and its parameters is valid, thus providing a one-step test of model adequacy. With the npde library, we provide an easy to use package for the computation of npde, thus making it easier to evaluate nonlinear mixed-effect models.

References:

- [1] F. Mentré and S. Escolano. Prediction discrepancies for the evaluation of nonlinear mixed-effects models. *J Pharmacokinet Biopharm* 33:345-67 (2006).
- [2] K. Brendel, E. Comets, C. Laffont, C. Laveille, and F. Mentré. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharm Res* 23:2036-49 (2006).
- [3] N. Holford, The Visual Predictive Check: superiority to standard diagnostic (Rorschach) plots, 14th meeting of the Population Approach Group in Europe, Pamplona, Spain (2005) Abstr 738.
- [4] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2006).