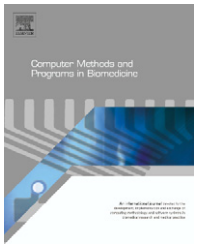
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 90 (2008) 154–166

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: The npde add-on package for R

Emmanuelle Comets[a,b,*], *Karl Brendel*[c], *France Mentré*[a,b,d]

[a] *INSERM, U738, Paris, France*
[b] *Université Paris 7, UFR de Médecine, Paris, France*
[c] *Institut de Recherches Internationales Servier, Courbevoie, France*
[d] *AP-HP, Hôpital Bichat, UF de Biostatistiques, Paris, France*

ARTICLE INFO

Article history:
Received 3 August 2007
Received in revised form
14 November 2007
Accepted 3 December 2007

Keywords:
Population pharmacokinetics
Population pharmacodynamics
Model evaluation
Nonlinear mixed-effects model
npde

ABSTRACT

Pharmacokinetic/pharmacodynamic data are often analysed using nonlinear mixed-effect models, and model evaluation should be an important part of the analysis. Recently, normalised prediction distribution errors (npde) have been proposed as a model evaluation tool. In this paper, we describe an add-on package for the open source statistical package R, designed to compute npde. npde take into account the full predictive distribution of each individual observation and handle multiple observations within subjects. Under the null hypothesis that the model under scrutiny describes the validation dataset, npde should follow the standard normal distribution. Simulations need to be performed before hand, using for example the software used for model estimation. We illustrate the use of the package with two simulated datasets, one under the true model and one with different parameter values, to show how npde can be used to evaluate models. Model estimation and data simulation were performed using NONMEM version 5.1.

1. Introduction

The analysis of longitudinal data is prominent in pharmacokinetic (PK) and pharmacodynamic (PD) studies, especially during drug development [1]. Nonlinear mixed-effect models are increasingly used as they are able to represent complex nonlinear processes and to describe both between and within subject variability. The evaluation of these models is gaining importance as the field of their application widens, ranging from dosage recommendation to clinical trial simulations [2]. Following the definition of Yano et al. [2]: "the goal of model evaluation is objective assessment of the predictive ability of a model for domain-specific quantities of interest, or to determine whether the model deficiencies (the final model is never the 'true model') have a noticeable effect in substantive inferences".

Despite the recommendations of drug agencies [3,4] stressing the importance of model evaluation, a recent survey based on all published PK and/or PD analyses over the period of 2002–2004 shows that it is infrequently reported and often inadequately performed [5]. One possible explanation is the lack of consensus concerning a proper evaluation method. Following the development of linearisation-based approaches for the estimation of parameters in nonlinear mixed-effect models, standardised prediction errors [6] have been widely

* *Corresponding author at*: INSERM U738, Université Paris 7, UFR de Médecine, site Bichat, 16 rue Henri Huchard, 75 018 Paris, France. Tel.: +33 1 44 85 62 77; fax: +33 1 44 85 62 80.
E-mail address: emmanuelle.comets@bichat.inserm.fr (E. Comets).
0169-2607/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2007.12.002

used as diagnostic tools, not the least because they were computed in the main software used in population PKPD analyses, NONMEM[7], where they are reported under the name weighted residuals (WRES). However, because of the linearisation involved in their computation there is no adequate test statistic. In 1998, Mesnil et al. proposed prediction discrepancies, which were easily computed due to the discrete nature of the non-parametric distribution estimated, to validate a PK model for mizolastine [8]. Prediction discrepancies (pd) are defined as the percentile of an observation in the 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 (the so-called plug-in approach [9]). By construction pd follow a uniform distribution over [0,1], providing a test. In the Bayesian literature this idea of using the whole predictive distribution for model evaluation has been proposed by Gelfand et al. [10] and is also discussed by Gelman et al. [11]. Yano et al. extended this notion in a non-Bayesian framework, proposing the approach known as Posterior Predictive Check (PPC) [2], while Holford advocated a more visual approach under the name Visual Predictive Check (VPC) [12]. Mentré and Escolano [13] discuss how prediction discrepancies relate to one of the three forms of PPC described by Yano. For non-discrete distributions, Mentré and Escolano proposed to compute prediction discrepancies by Monte–Carlo integration [13,14]. In their original version, pd however did not take into account the fact that subjects usually contributes several measurements which induces correlations between pd, leading to increased type I error. This was improved in a further work, and the uncorrelated and normalised version of pd was termed normalised prediction distribution errors (npde) [15]. npde have better properties than WRES, and can also be used to evaluate covariate models [16]. They can be used for internal or external evaluation, depending on whether they are computed on the dataset used to build the model (internal evaluation) or on an external dataset.

The computation of the npde however requires some programming. We therefore developed an add-on package, npde, for R, the open source language and environment for statistical computing and graphics [17], to enable easy computation of the npde [18]. Other packages such as Xpose [19], for diagnostic and exploration, and PFIM [20,21], for the evaluation and optimisation of population designs, have been developed in R for the analysis of population PK and/or PD studies. Xpose is very useful as an aid for model assessment and run management for studies performed with the NONMEM software [7], widely used in this field but with next to no plotting capabilities, so that R was a good choice of language for the implementation of npde.

In Section 2 , we briefly recall how npde are computed. In Section 3 we describe the main features and usage of the package. In Section 4 we illustrate the use of the package with two simulated examples. The examples are simulated based on the well known dataset theophylline, available both in R and NONMEM: the first (V_{true}) is simulated with the model used for the evaluation, while the second (V_{false}) is simulated assuming a different set of parameters, and we show how npde can be used to reject the model for V_{false} but not for V_{true}.

2. Computational method and theory

2.1. Models and notations

Let B denotes a building (or learning) dataset and V a validation dataset (V can be the same as B for internal evaluation). B is used to build a population model called M^B. Evaluation methods compare the predictions obtained by M^B, using the design of V, to the observations in V. V can be the learning dataset B (internal evaluation) or a different dataset (external evaluation). The null hypothesis (H_0) is that data in the validation dataset V can be described by model M^B.

Let i denote the ith individual ($i = 1, \ldots, N$) and j the jth measurement in an individual ($j = 1, \ldots, n_i$, where n_i is the number of observations for subject i). Let n_{tot} denote the total number of observations ($n_{tot} = \sum_i n_i$). Let Y_i be the n_i-vector of observations observed in individual i. Let the function f denote the nonlinear structural model. f can represent for instance the PK model. The statistical model for the observation y_{ij} in patient i at time t_{ij}, is given by:

$$y_{ij} = f(t_{ij}, \theta_i) + \epsilon_{ij} \quad (1)$$

where θ_i is the vector of the individual parameters and ϵ_{ij} is the residual error, which is assumed to be normal, with zero mean. The variance of ϵ_{ij} may depend on the predicted concentrations $f(t_{ij}, \theta_i)$ through a (known) variance model. Let σ denote the vector of unknown parameters of this variance model.

In PKPD studies for instance, it is frequently assumed that the variance of the error follows a combined error model:

$$\mathrm{var}(\epsilon_{ij}) = \sigma_{inter}^2 + \sigma_{slope}^2 f(t_{ij}, \theta_i)^2 \quad (2)$$

where σ_{inter} and σ_{slope} are two parameters characterising the variance. In this case, $\sigma = (\sigma_{inter}, \sigma_{slope})'$. This combined variance model covers the case of an homoscedastic variance error model, where $\sigma_{slope} = 0$, and the case of a constant coefficient of variation error model when $\sigma_{inter} = 0$.

Another usual assumption in PKPD analyses is that the distribution of the individual parameters θ_i follows a normal distribution, or a log-normal distribution, as in:

$$\theta_i = h(\mu, X_i)\, e^{\eta_i} \quad (3)$$

where μ is the population vector of the parameters, X_i a vector of covariates, h is a function giving the expected value of the parameters depending on the covariates, and η_i represents the vector of random effects in individual i. η_i usually follows a normal distribution $\mathcal{N}(0, \Omega)$, where Ω is the variance-covariance matrix of the random effects, but other parametric or non-parametric assumptions can be used for the distribution of the random effects, as in the first paper proposing prediction discrepancies in the context of non-parametric estimation [8]. Although npde were developed in the area of PK and PD analyses, they are a general way of evaluating mixed-effect models