Dose escalation studies: a comparison among Bayesian models

A. Russu[1], G. De Nicolao[1], I. Poggesi[2], R. Gomeni[2]

[1] Department of Computer Engineering and Systems Science, University of Pavia, Italy
[2] Clinical Pharmacology/Modelling & Simulation, GlaxoSmithKline, Verona, Italy

Introduction

In recent years there has been a growing interest in Bayesian methods applied to Phase I dose escalation studies[1,2,3]. Bayesian population models provide predictions and confidence intervals of dose-exposure curves in a straightforward manner, both for individuals already enrolled in the trial and for a new, previously untested subject. In contrast to traditional mixed effects models, parameter uncertainty is accounted for in a thorough way.

Objectives

In the present communication, the performance of four alternative Bayesian population models was evaluated on real and simulated datasets.

Three parametric methods and a novel approach based on population smoothing splines[4] were implemented. A full model comparison procedure was developed for the selection of the most appropriate model.

Methods

Data

Ten Phase I dose escalation studies and 300 simulated datasets (generated with the three parametric models described below) were analyzed.

Models

The dose-exposure relationship was explored using:

- a linear model in log-log scale (**LL**, in short),
- a power model (**P**),
- an E_{max} model (**E_{max}**),
- a nonparametric model based on population smoothing splines (**SS**).

Parametric models were estimated using WinBUGS 1.4.3, splines using R 2.8.0. In all cases, a Bayesian population approach was adopted for model estimation. Noninformative priors were chosen for population parameters and measurement error variance. Priors of inter-individual variances in the parametric models were automatically tuned through a preliminary two-stage fitting.

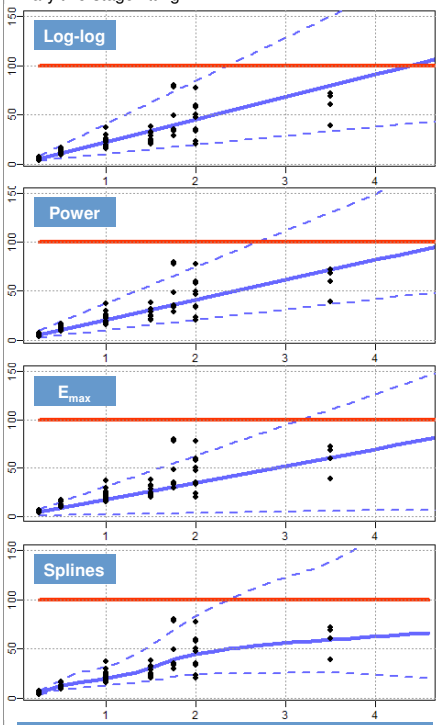

Figure 1: analysis of experimental data (C_{max} vs dose, study B in Table 1) with models LL, P, E_{max} and SS

Results – simulated data

All four models were fitted to data generated with the parametric models (Log-log/Power/E_{max} datasets in Figure 2 panel titles). Figure 2 shows boxplots for the weighted sum of squared residuals with respect to the true data (True WSSR), AIC, BIC and Cross-validatory Root Mean Square Error (CV RMSE).

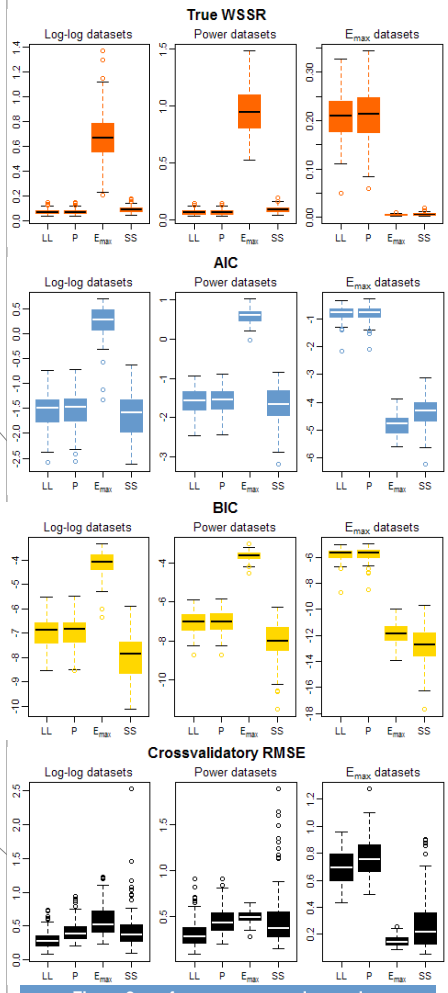

Figure 2: performance comparison using AIC, BIC, true WSSR and CV RMSE (top to bottom)

The proportion of correctly recognized models using the four performance metrics is shown in Figure 3 for data generated with LL, P, E_{max}, respectively. The lower proportions of correctly recognized models for the LL and, especially, the P case can be explained by the similarity of the two models (constant variance error for LL, constant CV error for P).

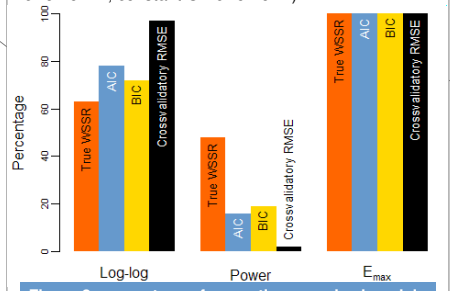

Figure 3: percentage of correctly recognized models using True WSSR, AIC, BIC, and CV RMSE

Results – experimental data

Model performances were evaluated on the ten experimental datasets (labelled A to J). For each dataset and estimation model, Table 1 reports:

- AIC (upper half of cell);
- Crossvalidatory RMSE (lower half).

Such choice emphasizes the contribution of both model complexity and predictive capability in assessing model performances.

Scores in bold denote the best method, according to the each of the two metrics. Smoothing splines, being structurally different from parametric models, were not considered for comparison purposes: their scores are reported as a reference for the other methods.

Dataset	Log-log	Power	E_{max}	Splines
A	**1,007**	1,114	1,716	**0,990**
	0,180	0,524	0,373	**0,262**
B	1,446	**1,381**	2,198	**0,692**
	0,525	0,517	**0,460**	**2,574**
C	1,480	1,608	**1,460**	**0,869**
	0,513	0,332	**0,275**	**0,388**
D	**-1,078**	-1,042	0,178	**-1,213**
	0,538	0,503	**0,415**	**0,587**
E	**0,532**	0,726	0,863	**0,462**
	0,246	0,246	0,278	**0,256**
F	**4,320**	5,502	4,334	**4,153**
	2,270	1,848	**0,999**	**2,295**
G	**0,047**	2,614	3,509	**1,863**
	2,032	4,781	**0,790**	**0,572**
H	**2,291**	2,945	2,476	**2,238**
	0,445	1,873	0,746	**0,450**
I	1,238	2,175	**0,370**	**0,335**
	0,284	0,289	**0,241**	**0,306**
J	**0,956**	1,147	2,562	**0,960**
	0,289	**0,287**	0,394	**0,288**

Table 1: performance comparison using AIC (upper half of cell) and CV RMSE (lower half). Best scores of parametric models are shown in bold

Conclusions

- A thorough model comparison procedure was developed, based on model complexity criteria and crossvalidatory techniques. Applying several criteria represents a useful cross-check when one is faced with the problem of finding the most adequate model.
- The Log-log and Power models yielded comparable results, both in simulated and experimental scenarios.
- Population splines may represent an appealing first-try, especially in early escalation stages, when there is not enough information to support a specific parametric model. Then, the proposed model comparison procedure can be applied for the selection of the most appropriate model.
- The proposed approach proved to robustly handle a variety of experimental scenarios, thus overcoming possible misspecification problems. In the simulation benchmark, considering the similarity between LL and P, our model comparison procedure identified the correct parametric model in most cases.

References

[1] J. Whitehead et al., Easy-to-implement Bayesian methods for dose escalation studies in healthy volunteers, *Biostatistics* 2, 47–61, 2001.
[2] D.A. Berry et al., Adaptive Bayesian designs for dose-ranging drug trials, in Gatsonis C et al. (eds.), Case Studies in Bayesian Statistics V. New York: Springer-Verlag, 99–181, 2001.
[3] A. Russu et al., Dose escalation studies: a comparison between NONMEM and a novel Bayesian tool. Population Approach Group in Europe (PAGE) 2008 Meeting, Marseille, France, June 2008.
[4] G. Pillonetto et al., Fast algorithms for nonparametric population modeling of large data sets, *Automatica* 45, 173–179, 2009.