Development and performance of npde for the evaluation of time-to-event model

Marc Cerou [1,2,3]★, Marc Lavielle [4], Karl Brendel [3], Marylore Chenel [3], Emmanuelle Comets [1,2]

[1] *INSERM UMR1137 IAME & Univ Paris Diderot, Paris, France* [2] *CIC 1414 & Univ Rennes 1, Rennes, France*
[3] *Institut de Recherches Internationales Servier , Suresnes, France* [4] *Inria & Ecole Polytechnique, Palaiseau, France*

Objectives: To develop normalised prediction distribution errors (npde) for time-to-event data and to diagnose their performance by considering type I error and power, using simulation studies.

Introduction

- Non-linear mixed effect models increasingly used in the drug development process
  - biomarkers followed throughout the trial and linked with the primary outcome (survival)
  - best described using joint models, which reduce bias in the estimation of model parameters
- Model evaluation
  - defined as assessing the adequacy between the tested model M and the data V
  - important part of model development [1]
  - graphical and statistical methods available for continuous data, including VPC [2], npde [3]
  - few methods, mostly visual, for non-continuous data
- Objective: develop npde for the time-to-event component of joint models

Methods

Statistical models

- Let T_i be the observation of the outcome in subject i and C_i the indicator of censoring. If $C_i = 0$, T_i corresponds to the time of the event, if $C_i = 1$ T_i is the censoring time.
- Model for the time-to-event (TTE) outcome characterised by the instantaneous risk h

$$h(t|f(\theta_i,t)) = h_0(t) \times \exp(\beta f(\theta_i,t))$$

  - where h_0 is the baseline hazard
  - f: structural model (which can be non-linear) of the longitudinal outcome
    - individual parameters $\theta_i \sim \mathcal{D}(\mu,\Omega)$ for the subject i, with fixed effects μ and variance-covariance matrix Ω
  - joint model: β represents the strength of the link between the longitudinal outcome and the time to the event

⇒ Model to evaluate:

$$M = \{f, h, \psi = (\mu, \Omega, \beta)\}$$

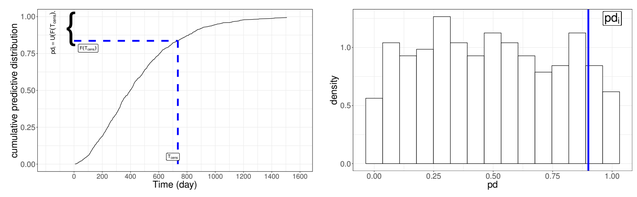
Model evaluation with npde

Construction

- npde are based on prediction discrepancies pd, defined as the quantile of an observation in its predictive distribution

$$pd_i = F_i(T_i) = \int^{T_i} p_i(t|\Psi)dt = \int^{T_i} \int p(t|\theta_i,\Psi)p(\theta_i|\Psi)d\theta_i dt$$

  - where F is approximated by Monte-Carlo simulations (K=1000)
  - for censored events, pd are imputed under the predictive distribution (similarly to continuous data below the LOQ[4])
    - extends to interval-censored TTE



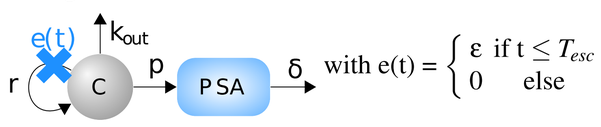
Test and Graphs

- $pd \sim U(0,1)$ and $npde = \phi^{-1}(pd) \sim \mathcal{N}(0,1)$
  - combined test on the distribution of npde [3]: Wilcoxon test for mean, Fisher test for variance, and Shapiro-Wilk test for normality adjusted with the correction of Bonferroni
- Visual diagnostics
  - QQplot with 95% confidence interval
  - histogram of the npde

Simulation

Motivating example: Desmée et al. 2015 [5]

- Metastatic castration-resistant prostate cancer
- Model developed using data from a phase III clinical trial
  - N=500 individuals with 735 days maximum follow-up
  - longitudinal observations of PSA (Prostate Specific Antigen) every 21 days (maximum of 35 measurements)
- Longitudinal model:



- Time-to-event model:

$$h(t|PSA(\theta,t)) = h_0(t)\exp(\beta \times PSA(\theta,t))$$

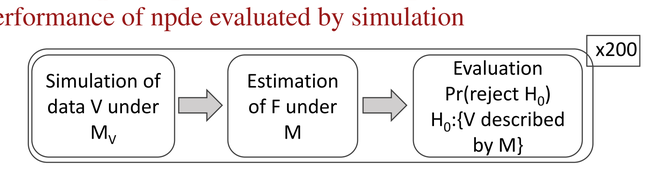
with h_0 characterized by a parametric Weibull model $h_0(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1}$

- Parameters estimated by the SAEM algorithm:

Parameter	Fixed effects	Transformation	IIV (ω)
r	0.05	log-normal	0.1
PSA_0	80	log-normal	0.6
ε	0.3	logit-normal	1.5
T_{esc}	140	log-normal	0.6
d	0.046	fixed	-
δ	0.23	fixed	-
k	1.52	-	0
λ	873	-	0
β	1.16×10^{-6}	-	0

Evaluating the performance of npde

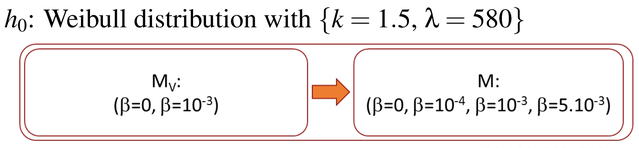
Performance of npde evaluated by simulation



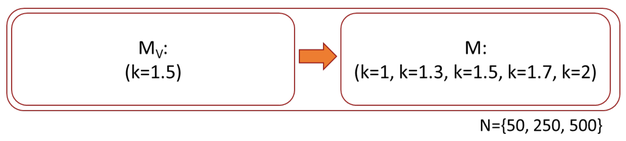
- M_V: model used to generate V
- **Type I error**: % of rejection of M under H_0 (M=M_V)
- **Power**: % of rejection of M under H_1 ($M \neq M_V$)
- Longitudinal model not evaluated and supposed to be correct

Simulation settings

- **Misspecification in the impact of PSA on survival** (β)
  - h_0: Weibull distribution with $\{k = 1.5, \lambda = 580\}$



- **Misspecification in the model of** h_0 (k)
  - h_0: Weibull distribution with $\{\lambda = 580\}$
  - same β for M_V and M ($\beta = 10^{-3}$)



Implementation

We used the statistical software R (version 3.2.3) and the package *mlxR* to simulate the data.

Results

Simulated Data

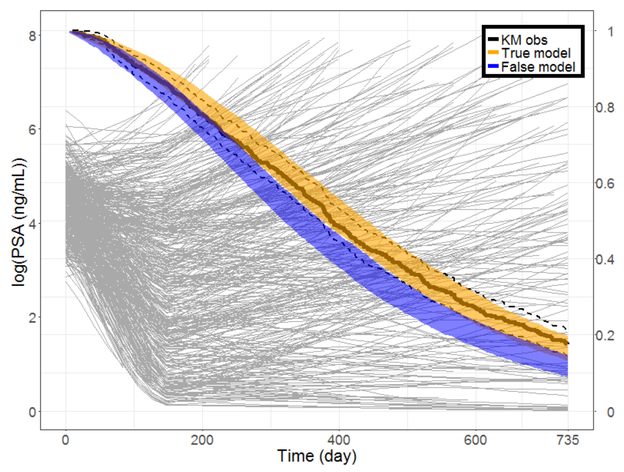


Figure 1: *Predicted PSA (grey) and survival Kaplan-Meier estimate of survival curve (with 95% CI in dashed line) for one simulated dataset under a Weibull model ($\{\beta = 10^{-3}, \lambda = 580, k = 1.5\}$). The yellow area (resp. blue) represents the 90% prediction interval of the survival function obtained from 1000 replicates simulated under the same (True) model (resp. False, $\beta = 5 \times 10^{-3}$).*

Using npde to evaluate a TTE model

- True model (left): under H_0, points remain in the prediction interval and the p-value is not significant
- False model (right): under H_1, most of the points are not in the prediction interval, and the distribution is shifted to the right as the model underestimates survival time, leading to p≤ 5%

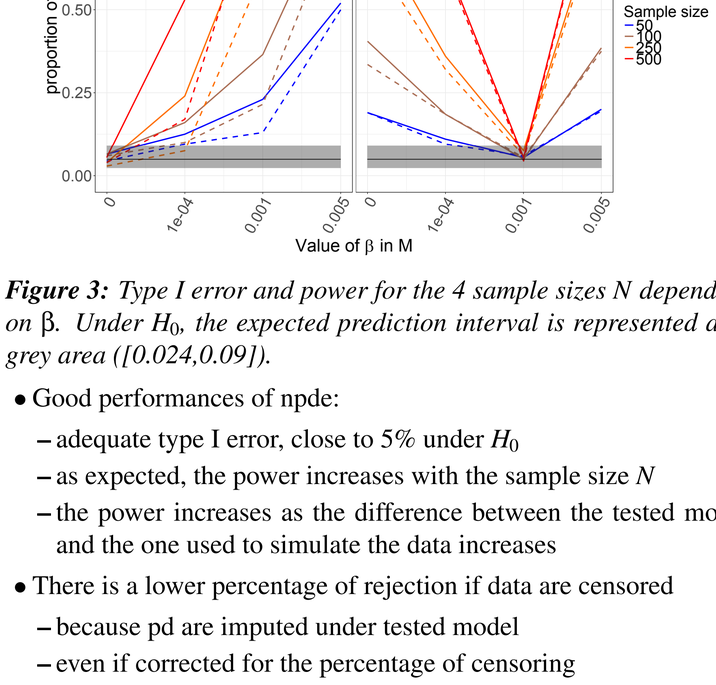


Figure 2: *Graphical and statistical diagnostics of the npde*

Misspecification in the impact of PSA on survival (β)

Figure 3: *Type I error and power for the 4 sample sizes N depending on β. Under H_0, the expected prediction interval is represented as a grey area ([0.024,0.09]).*

- Good performances of npde:
  - adequate type I error, close to 5% under H_0
  - as expected, the power increases with the sample size N
  - the power increases as the difference between the tested model and the one used to simulate the data increases
- There is a lower percentage of rejection if data are censored
  - because pd are imputed under tested model
  - even if corrected for the percentage of censoring

Misspecification in the model of h_0 (k)

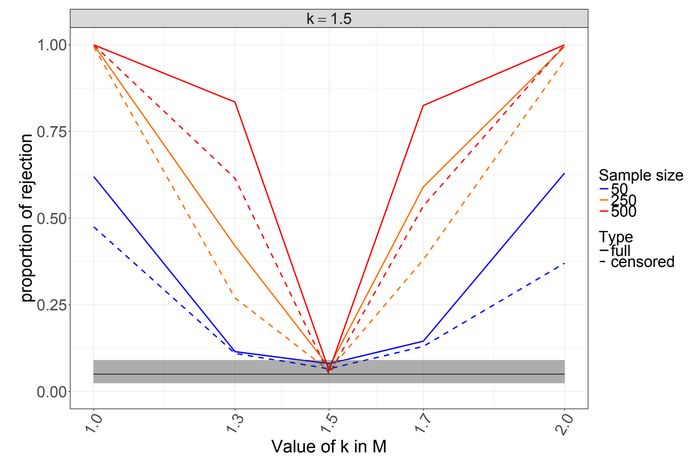


Figure 4: *Power of the npde to detect misspecification of h_0*

- npde able to pick up misspecification in h_0 with similar power when k is changed

Conclusion

- Development of npde for time-to-event data
  - can be extended to interval-censored TTE
- Good performance on simulated data
  - adequate type I error
  - power to detect model misspecifications in the survival model
    - misspecified link between the longitudinal marker and the outcome
    - misspecification in the shape of the survival curve
- Perspectives:
  - extension to joint evaluation when we consider time-to-event AND longitudinal observations
  - extension to joint modeling framework with repeated time-to-event and longitudinal observations

[1] Food and Drug Administration. *Guidance for Industry: Population Pharmacokinetics*. FDA, Rockville, Maryland, USA, 1999.

[2] M Karlsson and N Holford. A tutorial on visual predictive checks. *17th meeting of the Population Approach Group in Europe, Marseille, France*, page Abstr 1434, 2008.

[3] Karl Brendel, Emmanuelle Comets, Céline Laffont, and France Mentré. Evaluation of different tests based on observations for external model evaluation of population analyses. *Journal of Pharmacokinetics and Pharmacodynamics*, 37:49–65, 2010.

[4] T H T Nguyen, E Comets, and F Mentré. Prediction discrepancies (pd) for evaluation of models with data under limit of quantification. *Journal of Pharmacokinetics and Pharmacodynamics*, 39:499–518, 2012.

[5] Solène Desmée, France Mentré, Christine Veyrat-Follet, and Jérémie Guedj. Nonlinear mixed-effect models for prostate-specific antigen kinetics and link with survival in the context of metastatic prostate cancer: a comparison by simulation of two-stage and joint approaches. *The AAPS journal*, 17(3):691–699, 2015.

★ Presenting author email: marc.cerou@inserm.fr