



Use of mixture models in pharmacometric model-based analysis of confirmatory trials: part I - simulation study evaluating type I error and power of proof-of-concept trials

Adrien Tessier¹, Estelle Chasseloup², Mats O. Karlsson²

¹Pharmacometrics and Clinical Pharmacokinetics division, Servier, Suresnes, France

²Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

Objectives

Longitudinal analysis of response data can provide a valuable basis for precise estimation of drug effects and high power for detecting drug effects. However, concerns regarding Type 1 error control and bias in drug effect estimates often disqualifies such analyses from being the primary analysis method. We assessed the robustness of longitudinal analysis towards model misspecification in terms of type I error, power and estimation accuracy in the context of a PoC study and compared a standard and a new modelling approach, individual model averaging (IMA).

Methods

1. Dose response model for simulations

$\text{response}_{ij} = \text{baseline}_i + \text{placebo}_{ij} + \text{drg_slope}_i \times \text{dose}_i + \varepsilon_{ij}$
where $\text{placebo}_{ij} = \text{plcb_max}_i \times (1 - \exp(-\log(2)/\text{plcb_HL}_i \times \text{time}_{ij}))$
 $\theta_{\text{plcb_max}_i} = \theta_{\text{plcb_max}} \times (1 + \eta_{\text{plcb_max}_i})$
 $\theta_{p_i} = \theta_p \times \exp(\eta_{p_i})$ for $p = \text{baseline or drg_slope}$

2. Designs for simulations

Proof-of-concept trial design (POC): two parallel groups, dose = {0, 1}, 30 patients/group, response observed at time 0 (pre-dose), 1, 2 and 3
Larger sample size design (BIG): as POC but 120 patients/group

3. Nested models for the test of drug effect

Same model as simulations and 12 models included misspecifications of placebo or drug effect models were fitted to the simulated data

3.1. Standard approach (STD)

Drug effect parameters fixed to 0 (**base**) or estimated (**full**)

3.2. Individual model averaging approach (IMA)

Drug effect parameters estimated in both base & full models

mixture 1: IPRED = baseline + placebo + drug
mixture 2 or pre-dose: IPRED = baseline + placebo

base	full
\$MIX	\$MIX
P(1) = 0.5	P(1) = ARM _i × θ _x + (1 - ARM _i) × (1 - θ _x)
P(2) = 0.5	P(2) = 1 – P(1)
	ARM _i : randomization from the dataset (0 for placebo, 1 for active drug)

Individual model averaging method based on mixtures

$$OFV = \sum_{i=1}^N OFV_i = \sum_{i=1}^N -2 \log \left(\sum_{s=1}^S \exp(-OFV_{i,s}/2) \times P_{pop,s} \right) \quad [1]$$

s : sub-population (mixture) number ($s = 1, \dots, S$)

$OFV_{i,s}$: objective function value of individual i ($i = 1, \dots, N$) for the mixture s

$P_{pop,s}$: population probability of the mixture s

Simulation study process

Stochastic Simulation and Estimation

Performed with NONMEM 7.4.3 (FOCE) and PsN 4.8.8

base (or full) model specified for simulation of data for the POC/BIG design

base: simulation model without drug effect
full: simulation model with drug effect

×500 base (or full) dataset

×500 estimations with the base model

×500 estimations with the full model

estimation with nested models according STD or IMA approach

ΔOFV_1 ΔOFV_2 ... ΔOFV_{500}

Fraction of $\Delta OFV > \text{significance level}$
=
Type I Error (or Power)

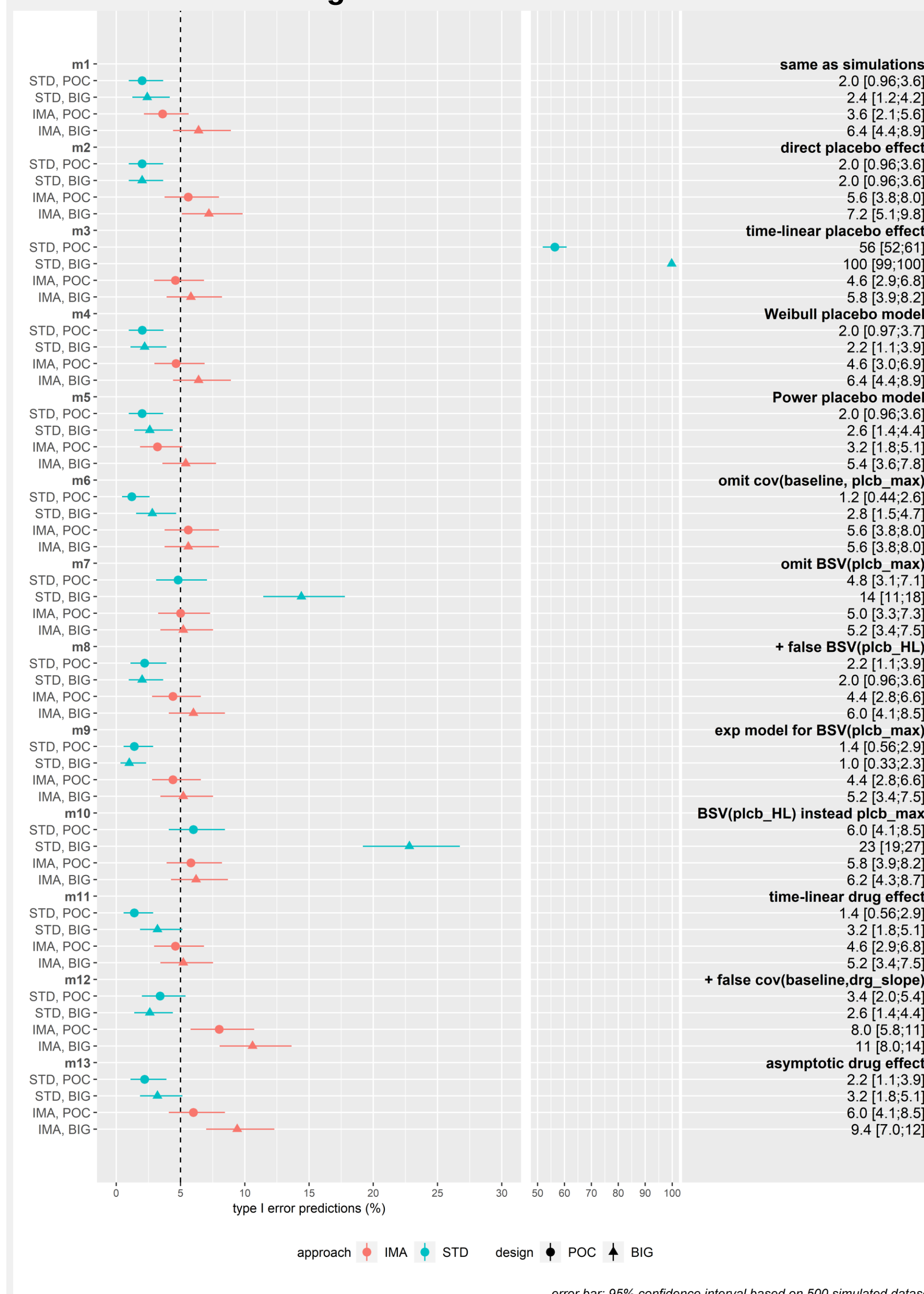
significance level = nominal cut-off value assuming chi-squared distribution of ΔOFV and type I error probability alpha of 5%

Conclusion

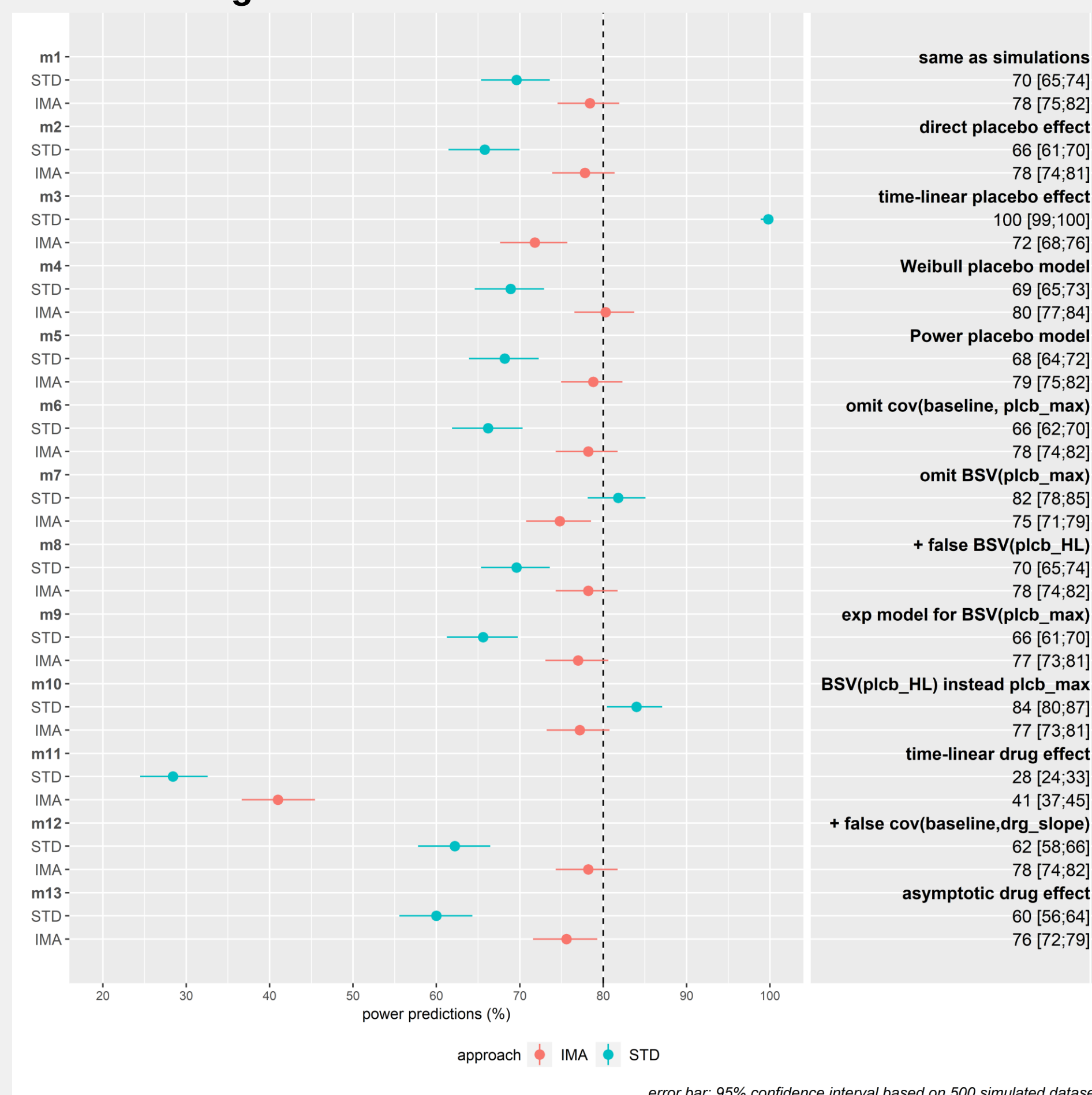
IMA was more robust towards model misspecifications with better type I error control compared to STD
IMA typically resulted in higher power than STD
IMA seems promising to evaluate the treatment effect in longitudinal data analysis with control of the type I error and accurate effect size estimates

Results

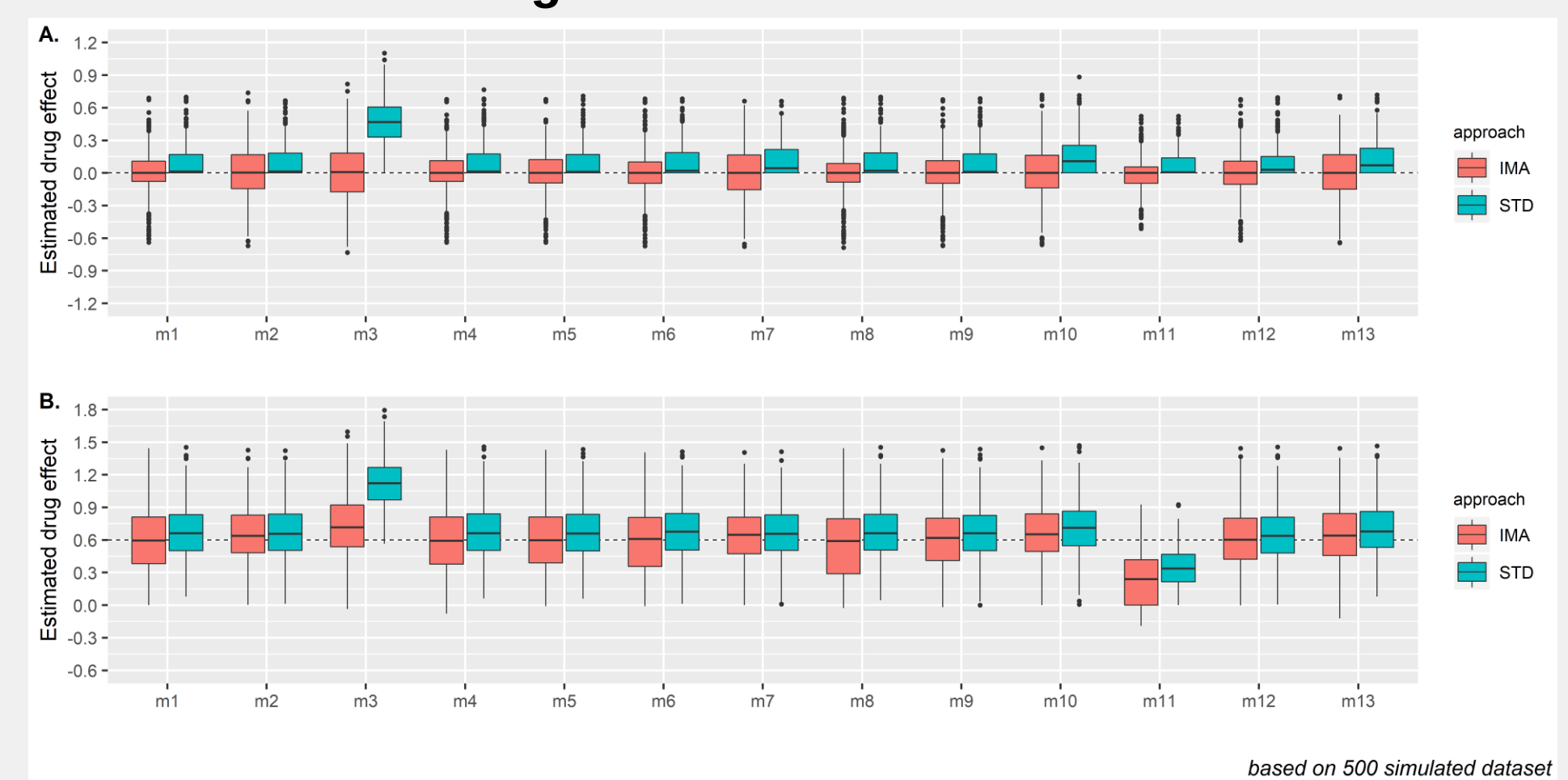
Type I error predictions for LRT at nominal cut-off values in POC and BIG designs



Power predictions for LRT at nominal cut-off values in POC design



Estimates of drug effect on base (A.) and full (B.) simulated dataset in POC design



drug effect = $\theta_{\text{drg_slope}} \times (\theta_x \times 2 - 1)$, with $\theta_x = 1$ for STD
simulated drug effect = 0 in base dataset
simulated drug effect = 0.6 in full dataset

STD

- type I error rates controlled for 3/13 (POC) and 2/13 models (BIG)
- type I error lower than pre-specified 5% in 8/13 models (BIG)
- inflated type I error in 3/13 models with placebo misspecifications (BIG)

IMA

- type I error rates controlled for 12/13 (POC) and 10/13 models (BIG)
- inflated type I error in 3/13 models with bigger drug models (BIG)
- higher power predictions than STD in 10/13 models (POC), but much higher type I error rates with STD for the 3 other models
- similar to more precise estimates of the drug effect with IMA compared to STD on both base and full simulated dataset

References:

[1] Karlsson, KC et al. Modeling Subpopulations with the \$MIXTURE Subroutine in NONMEM: Finding the Individual Probability of Belonging to a Subpopulation for the Use in Model Analysis and Improved Decision Making. AAPS J. 2009

For more results comparing STD and IMA approaches in terms of type I error using real data, see: Estelle Chasseloup et al. Poster I-68 (abstract 9149)