Model-based analyses for pivotal decisions, with an application to equivalence testing for biosimilars

B. Bieth*, F. Mentré**, G. Heimann*, I. Demin*,

B. Hamrén*, D. Renard*, S. Balser***

*** Modeling and Simulation, Novartis, Basel (Switzerland)**

**** UMR 738, INSERM and Universite Paris Diderot, Paris (France)**

***** Sandoz Biopharmaceutical, Holzkirchen (Germany)**

NOVARTIS

M&S
Modeling & Simulation

Introduction

- Problem.
  - **Can the size of the Phase III study be reduced by optimizing the choice of analysis methodology, as compared to using the default approach with the primary clinical endpoint at end of the study?**
- Innovative model-based analysis
  - Developed to reduce the required number of patients
  - The strict regulatory standards for phase III maintained as much as possible.
  - Use of a fully pre-specified Non-Linear Mixed Effects (NLME) analysis
  - This is not how NLME analyses are routinely used (*ad-hoc* exploratory analyses).

Biosimilar equivalence in Rheumatoid Arthritis (RA)

- In RA, a study would typically have 24-weeks duration and aim to show equivalence of ACR20 responder rates at week 24.
- The equivalence testing includes hypothesis testing based on differences in proportions of ACR20 responder rates at the end-point (week 24).
  - The classical equivalence test only uses data collected at week 24 and estimates responder rate in a traditional manner (use of binomial distribution, #success/#patients).
  - Equivalence is inferred when the entire confidence interval for the treatment effect falls within the equivalence testing.

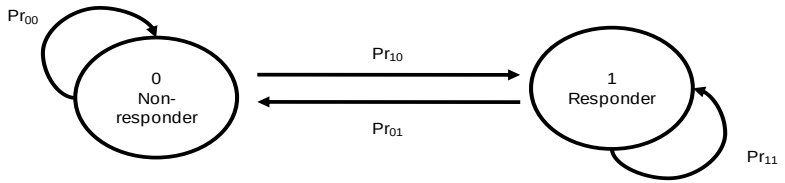
Longitudinal model-based test (1)

Pre-specified candidate models to capture longitudinal data

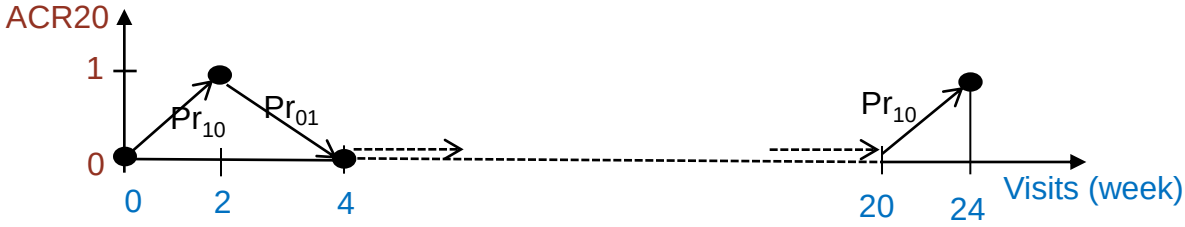
Model averaging

Equivalence testing

NLME models to describe transition probabilities using the Markov assumption linking entire longitudinal data.



Lacroix BD et al, Clin Pharmacol Ther 2009, 86: 387-395.



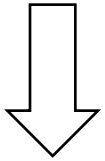
The estimate of the response rate at week 24 is obtained from the transition probabilities.

$$\begin{pmatrix} P_{ACR20=0} = 1 - P_{ACR20=1} \\ P_{ACR20=1} \end{pmatrix} = \prod_{k} \begin{pmatrix} 1-\Pr_{10}(t_k) & 1-\Pr_{11}(t_k) \\ \Pr_{10}(t_k) & \Pr_{11}(t_k) \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

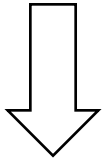
NOVARTIS

Longitudinal model-based test (2)

Pre-specified candidate models to capture longitudinal data



Model averaging



Equivalence testing

Markov model is specified for two independent transition probabilities e.g.

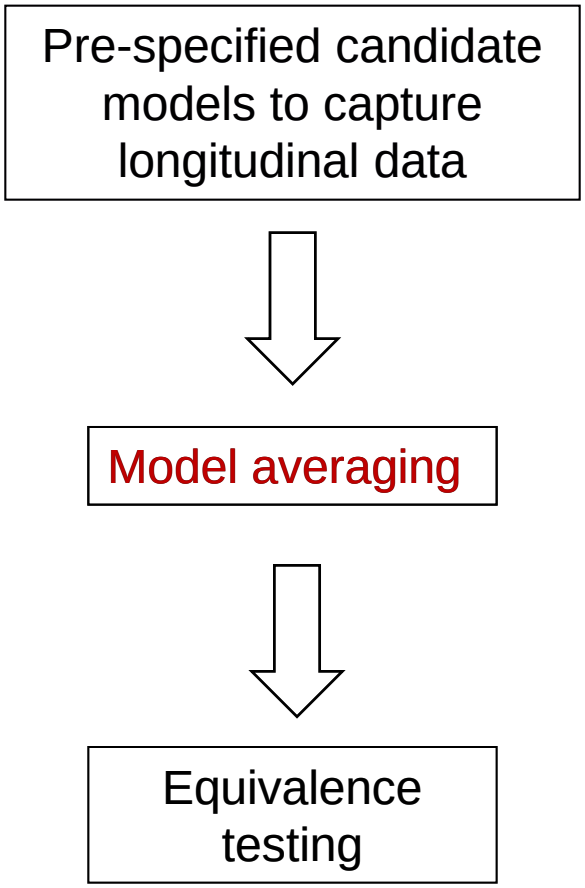
$$\log\left(\frac{\Pr_{10i}(t_k)}{1-\Pr_{10i}(t_k)}\right) = \beta_0 + \beta_1 TRT_i + (\beta_2 + \beta_3 TRT_i)\log(t_k) + \eta_{10i}$$

$$\log\left(\frac{\Pr_{11i}(t_k)}{1-\Pr_{11i}(t_k)}\right) = \beta_4 + \beta_5 TRT_i + (\beta_6 + \beta_7 TRT_i)\log(t_k) + \eta_{11i}$$

i= subject, k=occasion

- Individual response probability derived using previous formula
- Population response rate derived by integrating out the random effects (η_{10}, η_{11})
- Models differ by time dependency ($\log(t_K)$)

Longitudinal model-based test (3)



Key concept: model averaging

- Used to estimate the responder rates at week 24 by combining results from the different candidate models.
- Point estimate = weighted average of the individual model estimates
- Weights are based on BIC
- Bigger weights for models that fit the data better.

NOVARTIS

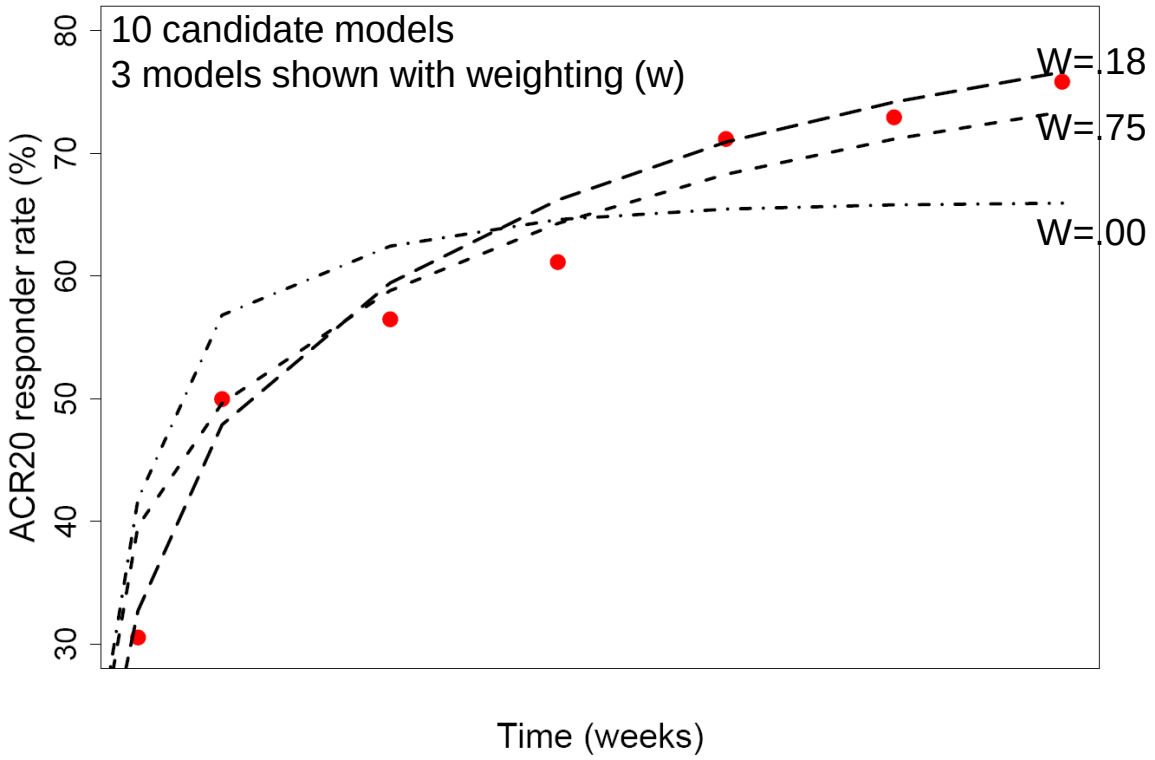
Longitudinal model-based test (4)

Pre-specified candidate models to capture longitudinal data

Model averaging

Equivalence testing

Example: model averaging



NOVARTIS

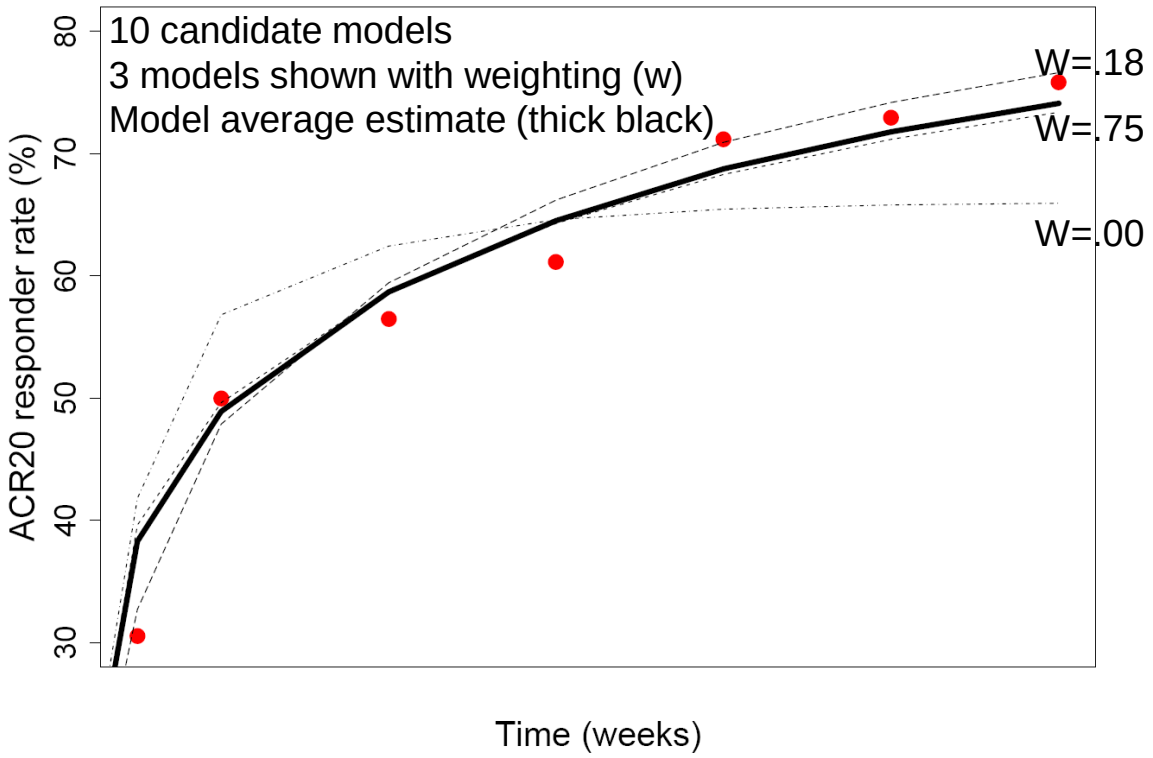
Longitudinal model-based test (5)

Pre-specified candidate models to capture longitudinal data

Model averaging

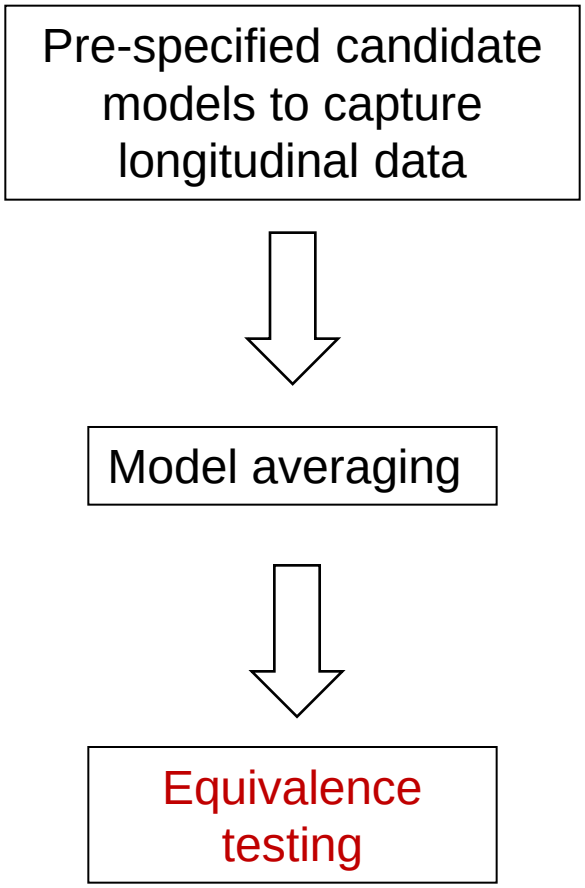
Equivalence testing

Example: model averaging



NOVARTIS

Longitudinal model-based test (6)

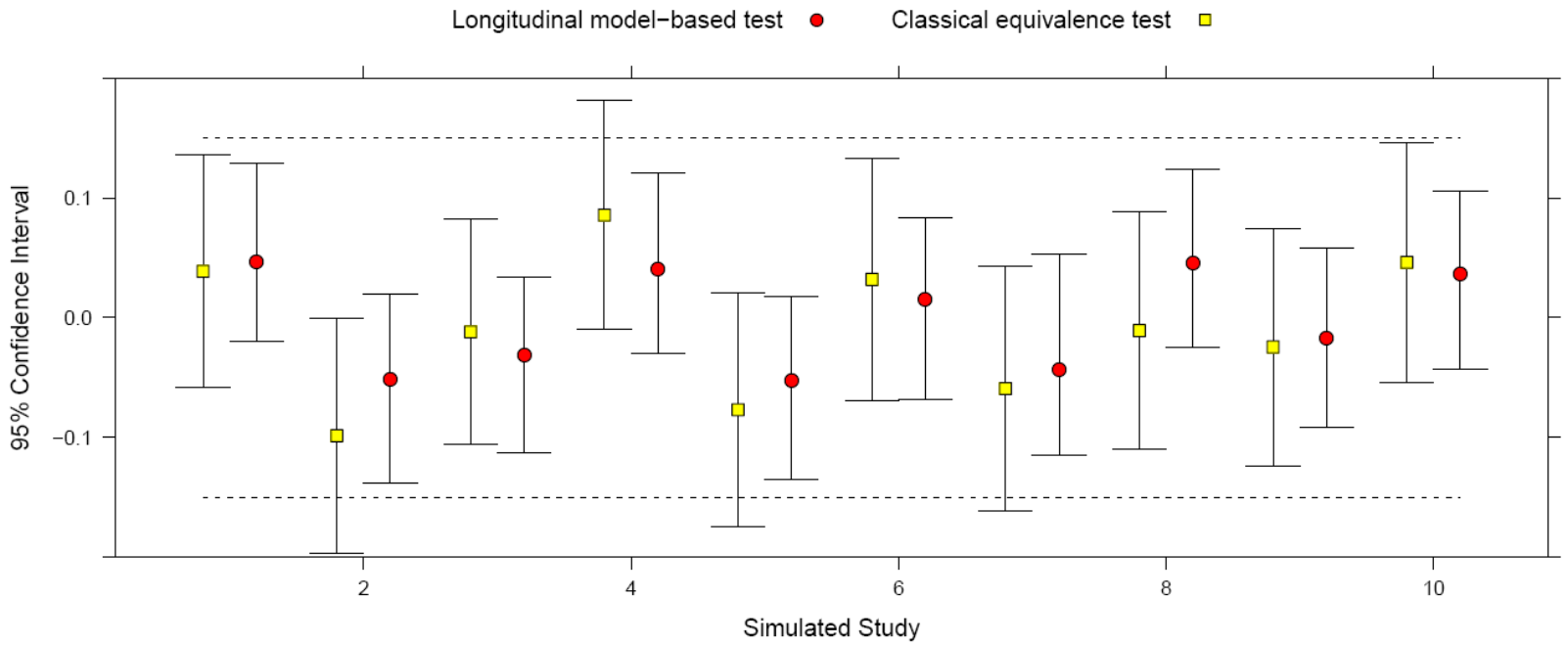


- Bootstrap is used to derive a confidence interval for the treatment difference at week 24.
- Bootstrap datasets are built by resampling over subjects.
- Confidence interval is compared with the equivalence margins for equivalence testing.

NOVARTIS

Model-based analysis vs. classical test

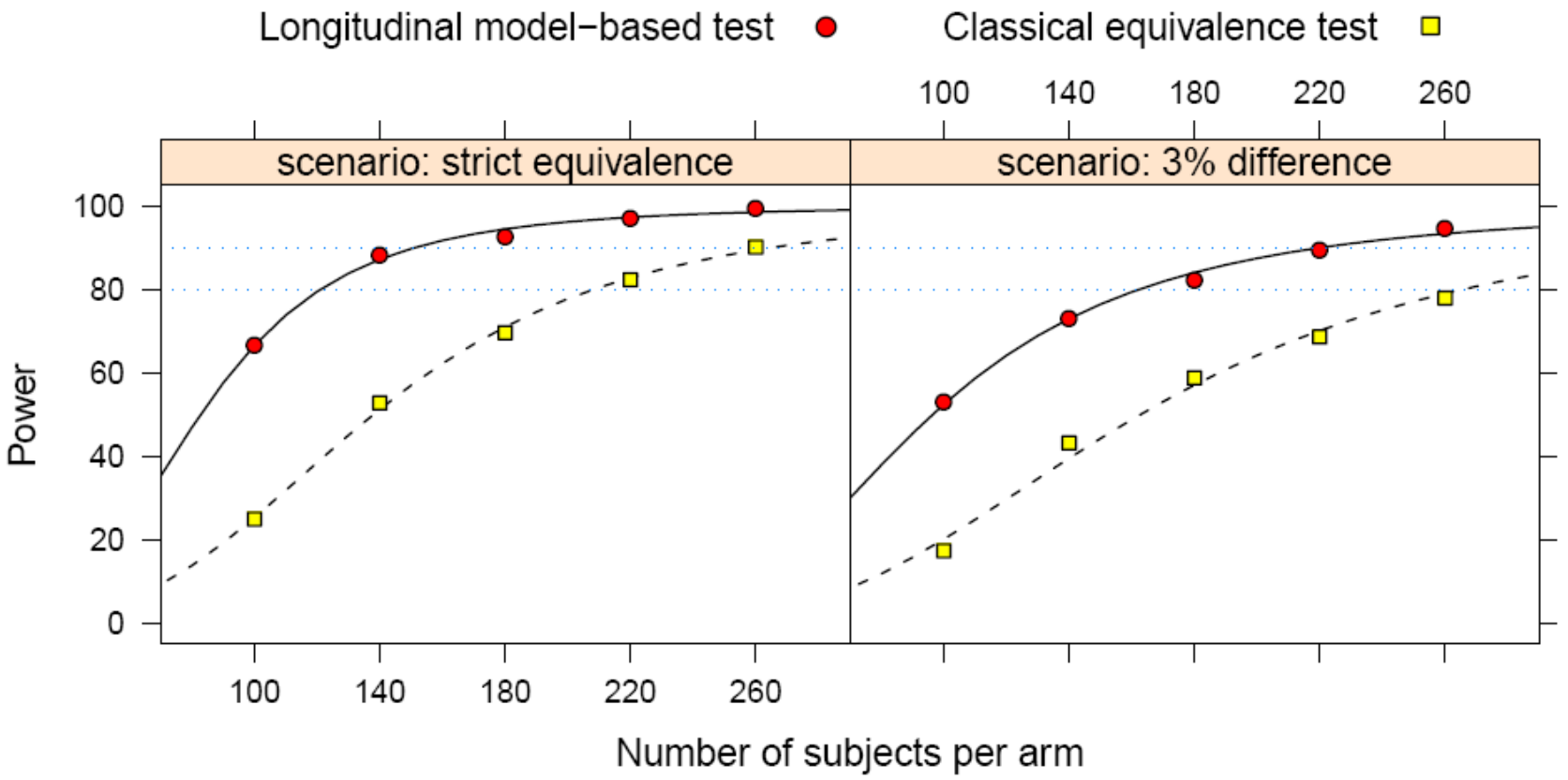
- Model-based approach does not change the nature of the comparability testing.



NOVARTIS

Simulation results

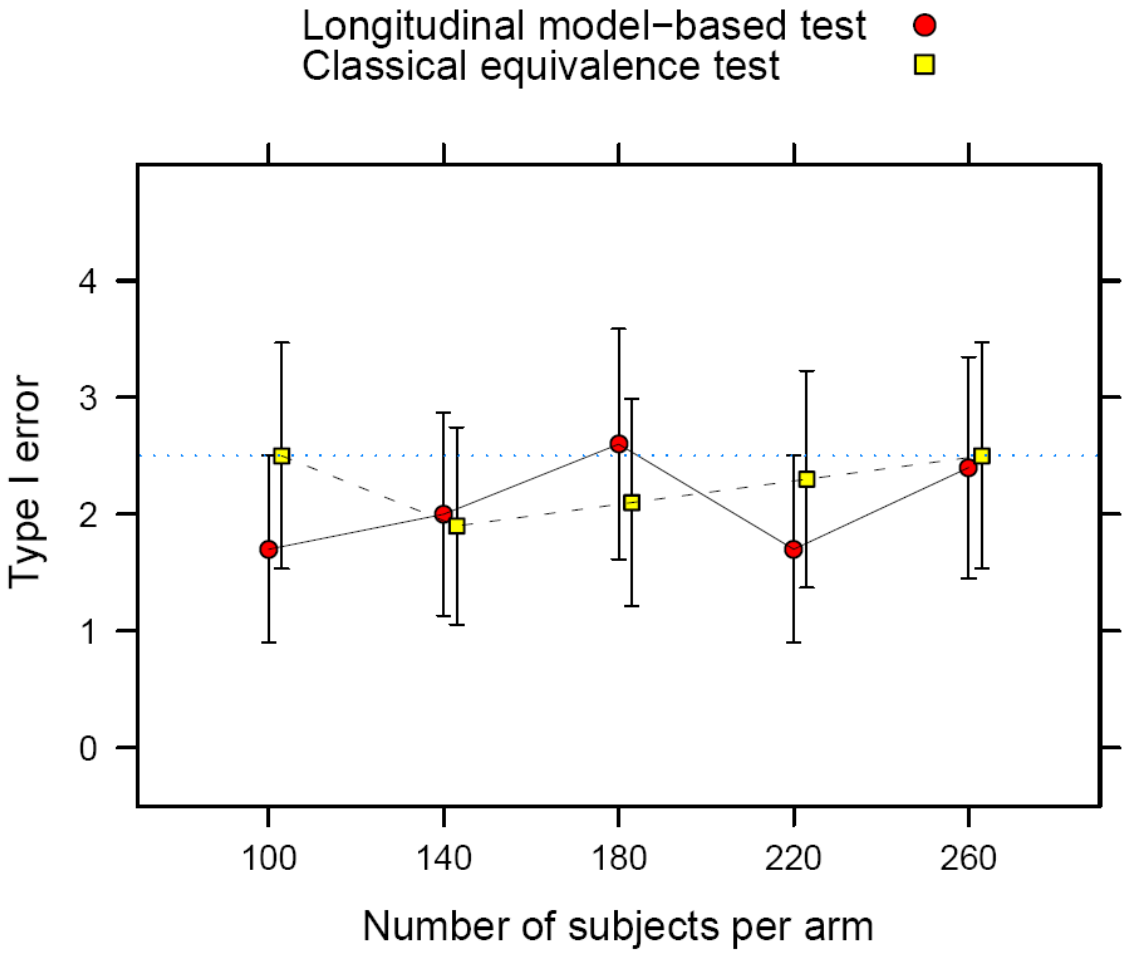
Required number of patients reduced by 40% compared to the classical test at power level of 80 & 90%



NOVARTIS

Simulation results

The type I error rate with the model-based test was close to the 2.5% nominal level.



NOVARTIS

Methodology Summary

- The model-based analysis uses all data collected to derive an estimate and its confidence interval of the treatment effect at the end of the study (week 24).
- Use of model-averaging to prevent against model misspecification.
- Number of patients reduced up to 40% using longitudinal model-based analysis – confirmed by additional simulation scenarios and sensitivity analysis.
- The general principles could be applied to other types of endpoints in RA (e.g. DAS28), in other therapeutic areas and in late stage clinical development, e.g. to analyze Phase 2 or 3 efficacy trials.

EMA Interaction

- Initial project feedback was negative.
- Overall encouraging feedback obtained at EMA/EFPIA workshop:
  - The absence of theoretical results to justify type I error control appears to be a critical concern deserving careful consideration.
- **How to get the acceptance for new methodologies from EMA?**
  - Ensure close interaction with EMA (Innovation Task Force).
  - Get regulatory acceptance via large simulation study because theoretical results seem out of reach.
  - Use model-based approach as supportive analysis in future studies.
  - Present and discuss the method with the scientific community.