EMA EFPIA workshop
Break-out session no. 4

Longitudinal model-based test as primary analysis in phase III

B. Bieth*, D. Renard*, I. Demin*, B. Hamrén*, G. Heimann*, Sigrid Balser**

* Modeling and Simulation, Novartis, Basel (Switzerland)

** Sandoz Biopharmaceutical, Holzkirchen (Germany)

Acknowledgement to Professors F. Mentre and J. Röhmel

BOS4 : Position statement and associated questions

- We propose to use a fully pre-specified longitudinal model-based test using all data as the primary analysis in Phase III:
  - Many indications require long term phase III studies.
  - Typically, the primary analysis estimates the treatment effect from data obtained at one time point of interest.
  - However the endpoint is frequently measured at many time points.
  - We propose an approach where the treatment effect at the time point of interest is estimated using **all** time points

- Does the regulatory agency agree that the proposed longitudinal model-based test is appropriate to be considered as primary analysis?
- If yes, what is the information and evidence that needs to be provided?
- If the answer is "no" at this point in time, what would it take to get acceptance for the proposed approach?

Background & Rationale

- The case study presented here is an example of using a longitudinal model-based approach in the framework of biosimilar equivalence testing in Rheumatoid Arthritis (RA).
- Regulatory guidelines require studies to assess equivalent efficacy of the biosimilar and the reference product.
  - In RA, a study would typically have 24-weeks duration and aim to show equivalence of ACR20 responder rates at week 24.
- The classical method for equivalence testing includes hypothesis testing based on differences in proportions at the end-point (week 24).
- The longitudinal model-based test uses all data collected to derive an estimate and its confidence interval of the treatment effect at week 24.
  - It does not change the nature of the comparability testing. The key difference is how the confidence interval is obtained.

Objective of the M&S work

- **Propose and evaluate, through clinical trial simulations, a fully pre-specified longitudinal model-based test for assessing biosimilarity of a test compound to a reference in RA.**

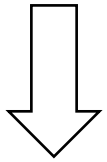
Methods / Summary

- **The equivalence testing problem is**

$$H_0 : | p_T - p_C | \geq \Delta \quad \text{versus} \quad H_1 : | p_T - p_C | < \Delta$$

with p_C the responder rate for the reference product and p_T for the biosimilar at week 24 and Δ=15% the pre-specified equivalence margin.

Cross-sectional data at week 24



Equivalence testing

The **classical equivalence** test only uses data collected at week 24 and estimates responder rate in a traditional manner (#success/#patients)

Longitudinal model-based test (1)

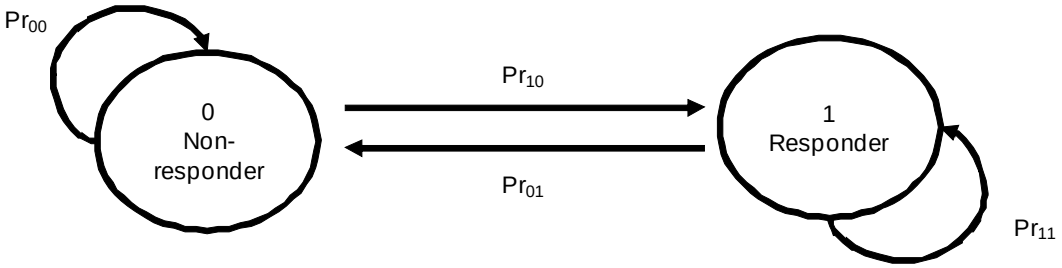
Pre-specified candidate models to capture longitudinal data

Model averaging

Equivalence testing

A number of logistic regression models are pre-specified to describe the transition probabilities as a function of time. For each model we rely on the Markov assumption to obtain the new estimate of the response rate at week 24 from the transition probabilities.

Transition probabilities



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

Note: first index refers to current visit, second index to previous visit.

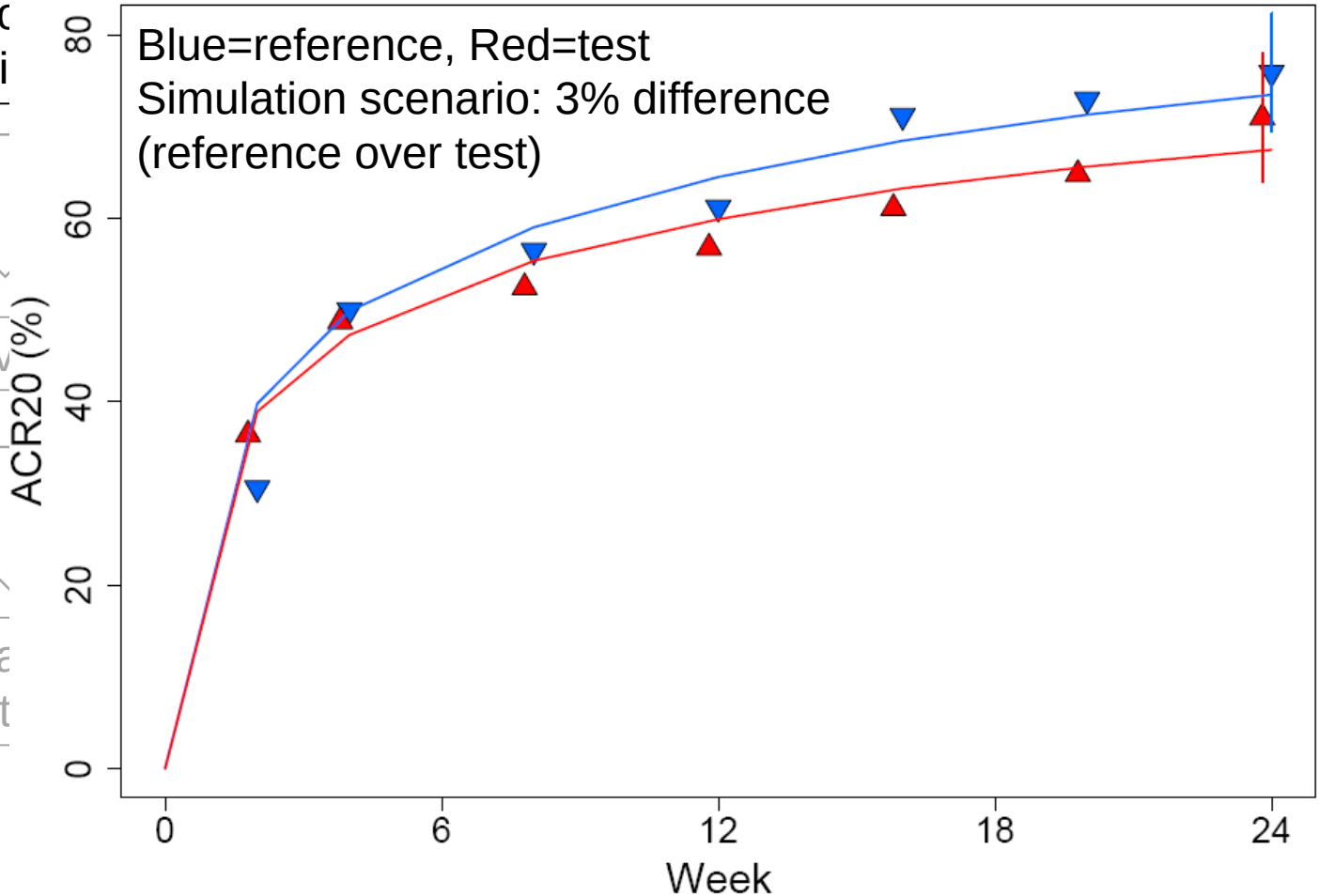
Longitudinal model-based test (1)

Pre-specified candidate models to longitudi

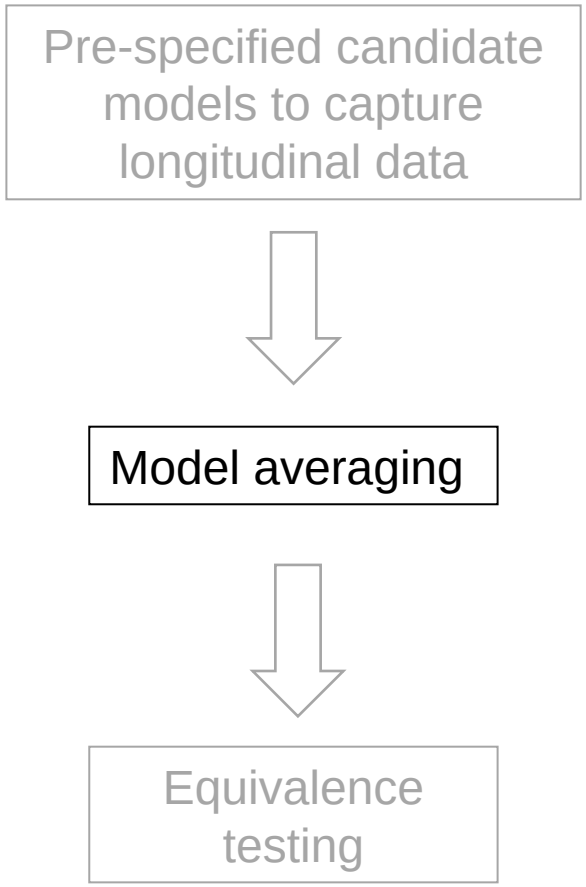
Model av

Equiva test

Example: one model



Longitudinal model-based test (2)



A number of logistic regression models are pre-specified to describe the transition probabilities as a function of time. For each model we rely on the Markov assumption to obtain the new estimate of the response rate at week 24 from the transition probabilities.

Model averaging is used to combine results from the different candidate models and estimate the responder rates at week 24 as weighted averages of the individual model estimates (bigger weights for models that fit the data well).

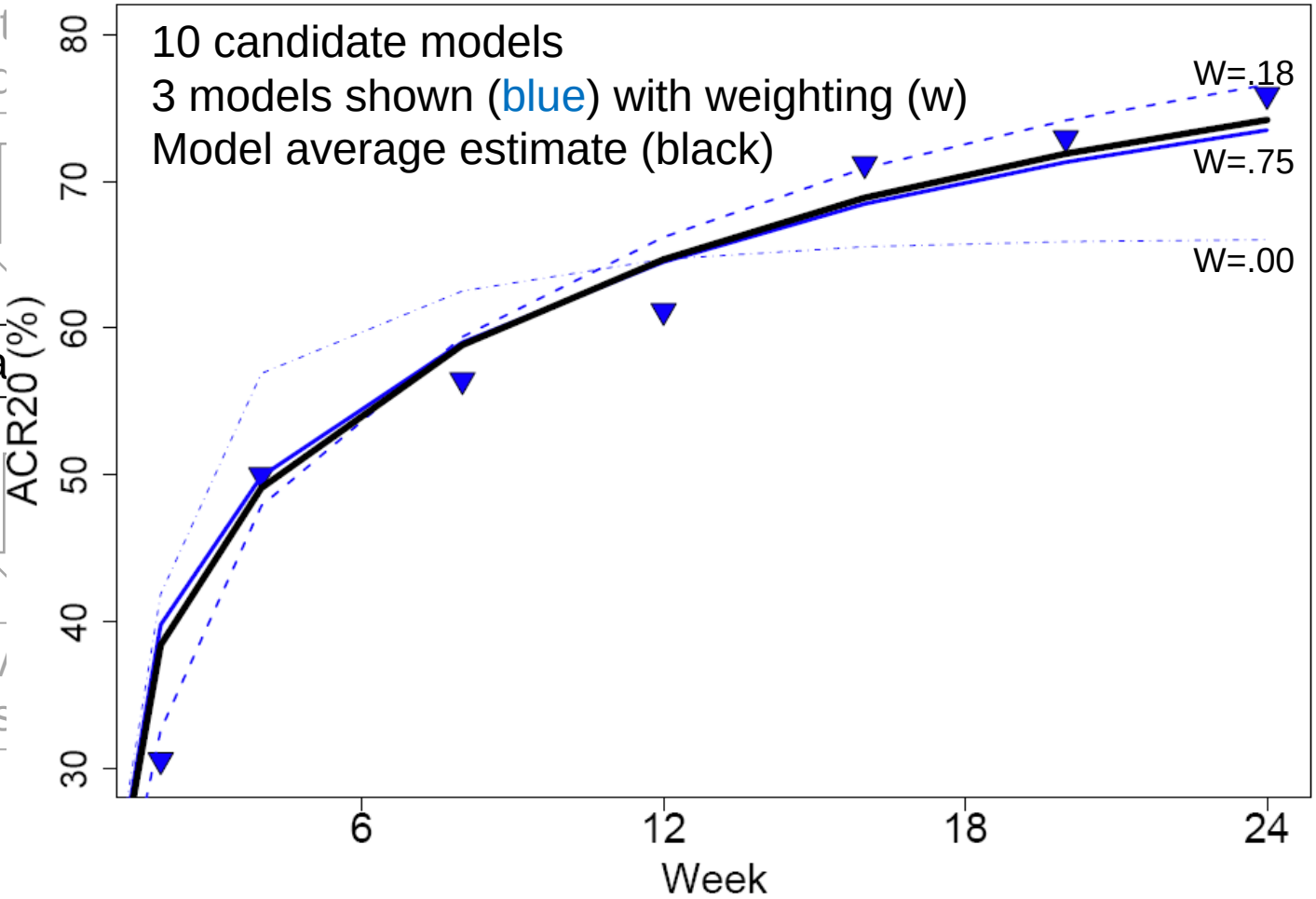
Longitudinal model-based test (2)

Pre-specified candidate models t
longitud

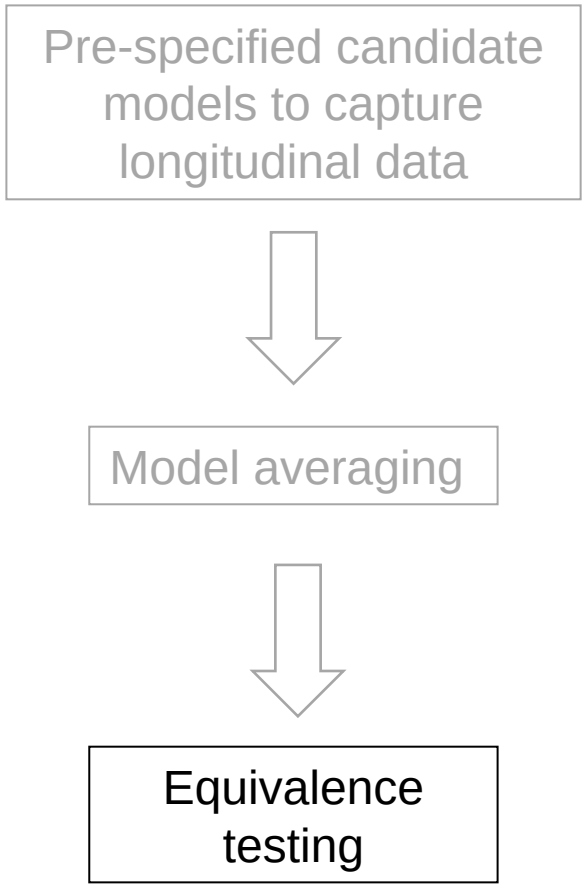
Model a

Equiv
tes

Example: model averaging



Longitudinal model-based test (3)



A number of logistic regression models are pre-specified to describe the transition probabilities as a function of time. For each model we rely on the Markov assumption to obtain the new estimate of the response rate at week 24 from the transition probabilities.

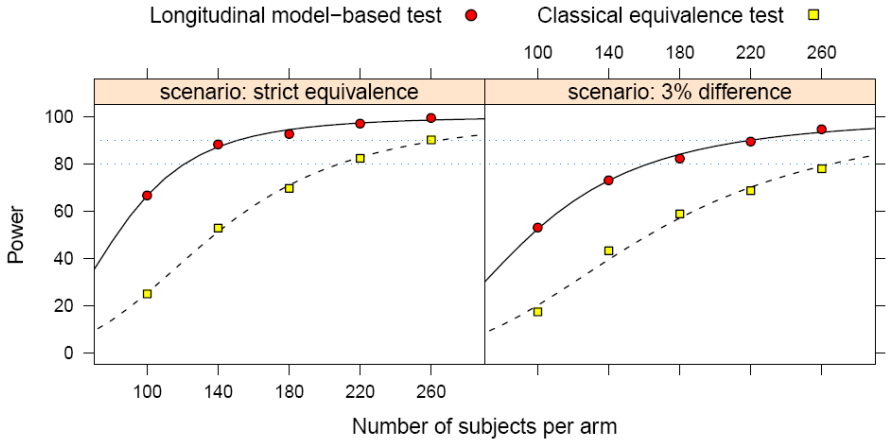
Model averaging is used to combine results from the different candidate models and estimate the responder rates at week 24 as weighted averages of the individual model estimates (bigger weights for models that fit the data well).

Bootstrap is used to derive a confidence interval for the treatment difference at week 24. It is compared with the equivalence margins for equivalence testing.

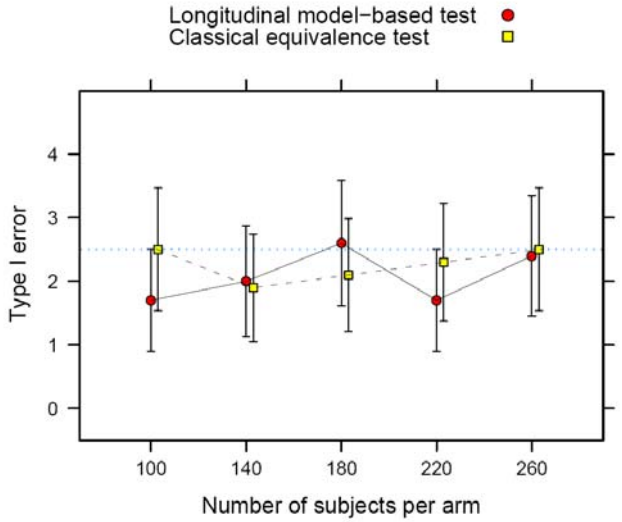
Significant gain in power using the longitudinal model-based test

Simulation results

The sample sizes using the model-based test were approximately **60%** of those using the classical test to reach target power levels of 80 and 90 %.



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



Conclusions

- The model-based test is centered on two key principles:
  - use of all data collected throughout the study period
  - use of model averaging to best characterize the treatment signal in this data and minimize model misspecification issues
- The proposed model-based test is shown via simulations to:
  - control the type I error rate
  - be much more efficient than the classical approach
- *It is proposed that such model based approaches should be considered suitable for the primary analysis of biosimilar trials*

Regulatory Feedback

- **Feedback from Scientific Advice Meeting** (only method description was submitted)

1. *"It is the opinion of CHMP that such an approach would not be able to overcome the problem of **type I error control**."*

2. *"According to current regulatory standards, the final approach chosen should consist of **one prespecified model**, which ideally **makes minimal assumptions**, and appropriate pre-specified **considerations on missing-data handling** should be made."*

M&S Novartis Questions to Regulator

1. Can the type I error concerns be addressed through simulations under a range of different scenarios?

2. Is the following sufficient to justify the proposed approach?
   - A single longitudinal model without assumptions on time course (saturated treatment-by-time mean structure) offers no efficiency gain over the end-point approach.
   - A single longitudinal model incorporating further assumptions on time course has increased risk of bias and poor performance due to model misspecification.
   - Missing data is not expected to be an issue in this application (per protocol population). However, what would be recommended in other applications (based on ITT population)?

BOS4 : Position statement and associated questions

- We propose to use a fully pre-specified longitudinal model-based test using all data as the primary analysis in Phase III:
  - Many indications require long term phase III studies.
  - Typically, the primary analysis estimates the treatment effect from data obtained at one time point of interest.
  - However the endpoint is frequently measured at many time points.
  - We propose an approach where the treatment effect at the time point of interest is estimated using the totality of time point

- Does the regulatory agency agree that the proposed longitudinal model-based test is appropriate to be considered as primary analysis?
- If yes, what is the information and evidence that needs to be provided?
- If the answer is "no" at this point in time, what would it take to get acceptance for the proposed approach?

Range of situations where this type of approach might be applied

- To assess biosimiliars where all preclinical (in vitro comparison) and clinical data (PK, biomarker) point to products being equivalent.
- To assess therapeutic equivalence of changes to formulation for a given dosage form e.g. topical, inhaled product , etc …
- To assess a new compound in an established therapeutic area where much is already known about the disease progression, placebo response and time course.
- To assess a new compound in a rare disease where patients are difficult to recruit (orphan indication).
- To assess a new compound in a new disease area (using extrapolation from other compounds).