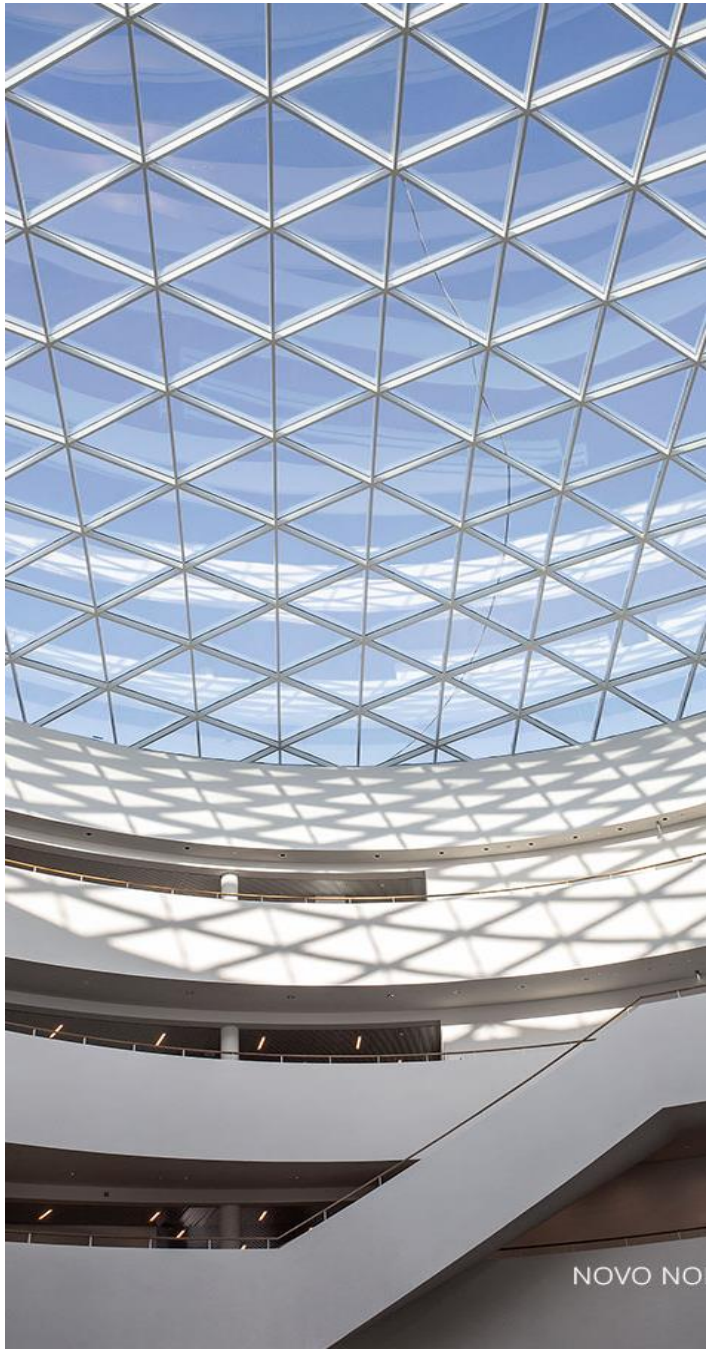


A confirmatory two-stage adaptive procedure for assessing average bioequivalence

Christian Pipper,

Biostatistics, Novo Nordisk A/S

April 20th , 2023

Disclaimer

The opinions expressed in these slides constitute the personal opinions of the author and not necessarily those of Novo Nordisk

Agenda

Setting and problem

Features of the design

Theoretical properties

Simulation performance

Some points to be challenged by

Settings and problem as they were presented to us

- **Context:** Initiated while working at Leo Pharma A/S with collaborators Kyle Raymond and Marie Louise Østerdal (Both statisticians at Leo)
- **Project:** Comparing two different administrations of an antibody treatment in atopic dermatitis patients.
- **Goal:** To claim bioequivalence of a new ”patient friendly” injection type compared to the standard one for admininstering the antibody treatment.
- **Challenge:** Uncertainty up front about the actual dosing/uptake of the drug with the new form of adminstration. The sneaky suspicion is that there might be under-dosing but how much is anyones guess
- **One possible solution:** we want to adapt to the suspected level of underdosing once we have some information on it→2 stage adaptive design with sample size reestimation

Initial wishes for design

- Approx 80 subjects for stage 1 balancing timelines and expected precision of pk endpoints (Cmax and AUC) in a parallel arm setting
- If stage 1 estimated geometric mean ratios (test versus reference) reflect more than 25% difference we give up on bioequivalence -- BUT we still want to know as much as possible about the magnitude of underdosing (that we suspect).
- An ”ordinary” 95% CI at the end of the trial consistent with any inferential decision about average bioequivalence to gauge the ”level of potential underdosing”
- As fast as possible →optimize the number of patients it takes in stage 2 to get an informative result

Our strategy:

Simultaneous inference on Cmax and AUC+no unnecessary retesting

Proposal: An adaptive TOST procedure-stage 1

- Let $\theta^{(l)}, \mathrm{l} = 1{,}2$ denote the targeted population summaries of Cmax and AUC

- We consider the following one-sided null hypotheses of non-bioequivalence:

$$H_0^-\colon \min\left(\theta^{(1)}, \theta^{(2)}\right) < -\Delta;\ \ H_0^+\colon \max\left(\theta^{(1)}, \theta^{(2)}\right) > \Delta$$

- Stagevise p-values for H_0^- equal those obtained for the smallest stagevise estimate
- Stagevise p-values for H_0^+ equal those obtained for the largest stagevise estimate

- Stage 1 p-values are evaluated separately for each hypothesis according to an efficiacy bound α_1 and binding futility bound α_0 (we have 0.5 in mind)

- A separate decision (stop for efficacy/futility or proceed) is made for each hypothesis

Proposal: An adaptive TOST procedure-stage 2

- If we proceed sample-size reestimation is made according to the scenario based on conditional power.
- Stagevise p-values are combined at stage 2 using a combination test.
- Critical value computed to ensure type 1 error control for each separate hypothesis
- Bioequivalence declared if both hypotheses are rejected.

Note: If we have already decided on a hypothesis at stage 1 it is not evaluated further after stage 2 (despite the fact that we could do it)

Type 1 error control

- For testing bioequivalence, that is: $H_0^- \cup H_0^+$
- Is ensured if type 1 error control is enforced when testing H_0^- and H_0^+ separately
- This is, in turn, ensured (asymptotically) if stagewise p-values are p-clud (asymptotically):

$$\lim_{n\to\infty} P\left(p_j \leq \alpha\right) \leq \alpha, for\ 0 \leq \alpha \leq 1$$

- Which you can show with some effort (The situation $\theta^{(1)} = \theta^{(2)}$ is non-trivial)

Overall confidence limits: Construction

- Based on overall p-values:

$$Q(p_1, p_2, \alpha_0, \alpha_1) = p_1 I(p_1 < \alpha_1 \; or \; p_1 \geq \alpha_0) + I(\alpha_1 \leq p_1 < \alpha_0)\{\alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 I(C(x,y) \leq C(p_1, p_2))dydx\}$$

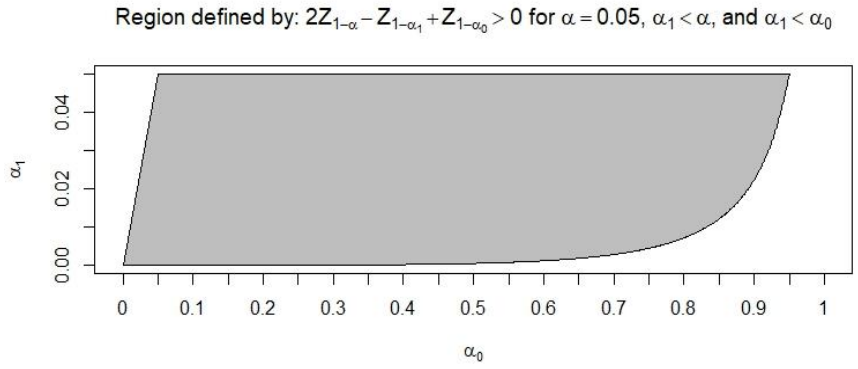
- Where *C* denotes the chosen combination function (f.i. inverse normal)

- Insert stagewise shifted p-values p_{l-} (H_0^-: $\min(\theta^{(1)}, \theta^{(2)}) \leq \delta$) and p_{l+} (H_0^+: $\max(\theta^{(1)}, \theta^{(2)}) \geq \delta$) into Q

- Solve $Q(p_{1-}, p_{2-}, \alpha_0, \alpha_1) = \alpha$ to obtain a lower 1-α lower confidence bound L for $\min(\theta^{(1)}, \theta^{(2)})$.
- Solve $Q(p_{1+}, p_{2+}, \alpha_0, \alpha_1) = \alpha$ to obtain a upper 1-α lower confidence bound U for $\max(\theta^{(1)}, \theta^{(2)})$.

Overall confidence limits: properties

- Confidence is ensured since shifted p-values are asymptotically p-clud when evaluated at the true parameter value
- Not up-front ensured that L<U←May be based on different data
- This can be ensured if loosely stated: stage 1 precision is large enough, does not vary too much between endpoints, and large values α_0 are avoided (how large depends in a complicated manner on efficacy bound and precision difference between endpoints)



One strategy for sample size reestimation

- Based on stage 1 estimates and ensuring overall power to reject:
  1. $H_0^- \cup H_0^+$
  2. H_0^-
  3. H_0^+
- Depending on the ”futility” decisions made at stage 1

- The basic focus is to minimize sample-size

- One out of many possible strategies (and maybe not the best) for sample-size re-estimation.

Simulation performance: some points learned

- Sensible confidence bounds (L<U) are not ensured per default. The restraints that are derived theoretically to ensure this do the job.
- The targeted degree of confidence is met in the simulations
- Evaluating both endpoints simultaneously gains power
- Minor gains in power are achieved by avoiding re-evaluations of decisions made at stage 1 (no re-testing of rejected/accepted null hypotheses.)

What literately not to like

- **The price of sensible confidence bounds**
  - With the outlined setup, it requires binding futility
  - And even restricts the simultaneous choice of efficacy and futility bound
  - These requirements are definitely controversial

- **Pooling of endpoint summaries**
  - If you aim for separate statements for each PK endpoint the ”win on all” approach presented here is not the way to go

Additional details

- An early version on this work is freely available on Arxiv: [2203.09182] Properties of a confirmatory two-stage adaptive procedure for assessing average bioequivalence (arxiv.org)