How the estimands framework affects the choice of non-inferiority margin

David Wright, AstraZeneca,

on behalf of the EIWG sub-team on estimands in non-inferiority trials

19 June 2024 PSI Annual Conference, Amsterdam, The Netherlands

Disclaimer

The views expressed by the presenter are not necessarily the views and practices of his employer, or of any of the EIWG member companies and organisations.

Time for a change in Mindset?

Agenda

- Motivation
- 2 Case studies
  - Case study 1: Defining a non-inferiority in weight management
  - Case study 2: Defining a non-inferiority margin in depression
- Key issues
- Questions for the panel discussion

Why is a meta-analysis needed?

- Confirm the comparator drug is superior to placebo/reference/standard of care
- Establish the size of “the effect” – you may ask which one!
- Different studies may have been conducted in different regions and/or under different conditions.
- You may also ask in what indication(s) has the drug demonstrated efficacy
- Which of these indications should you choose to demonstrate non-inferiority?
- Do you need to study all of those indications?

Motivation

- https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-development-guideline-non-inferiority-equivalence-comparisons-clinical-trials_en.pdf
  - Lists as one of the issues: difficulties in defining the non-inferiority margin – why?

- Virtually all historical studies that will be included in meta-analyses have not used the estimand framework and even if they had, a different estimand is often needed and used in a non-inferiority trial
  - So what estimand should be used when deriving the non-inferiority margin from historical studies?

Motivation

- Often key pieces of information are missing from the publications of both the individual studies and meta-analyses of these studies, for example:
  - Is the estimand you are interested available?
  - What are the relevant intercurrent events and were these considered in the choice of analyses conducted in these studies?
  - What was the frequency and timing of each intercurrent event in each study? Was it similar or different when studies are compared? If similar is this rate and timing expected to remain similar in the new study? If different what implications does that have to inclusion of certain studies in a meta-analyses?
  - How pragmatic can or should the choices of inclusion of studies in a meta-analysis be ? *e.g.* if some information is missing or if studies are of slightly different duration.
  - Constancy assumption

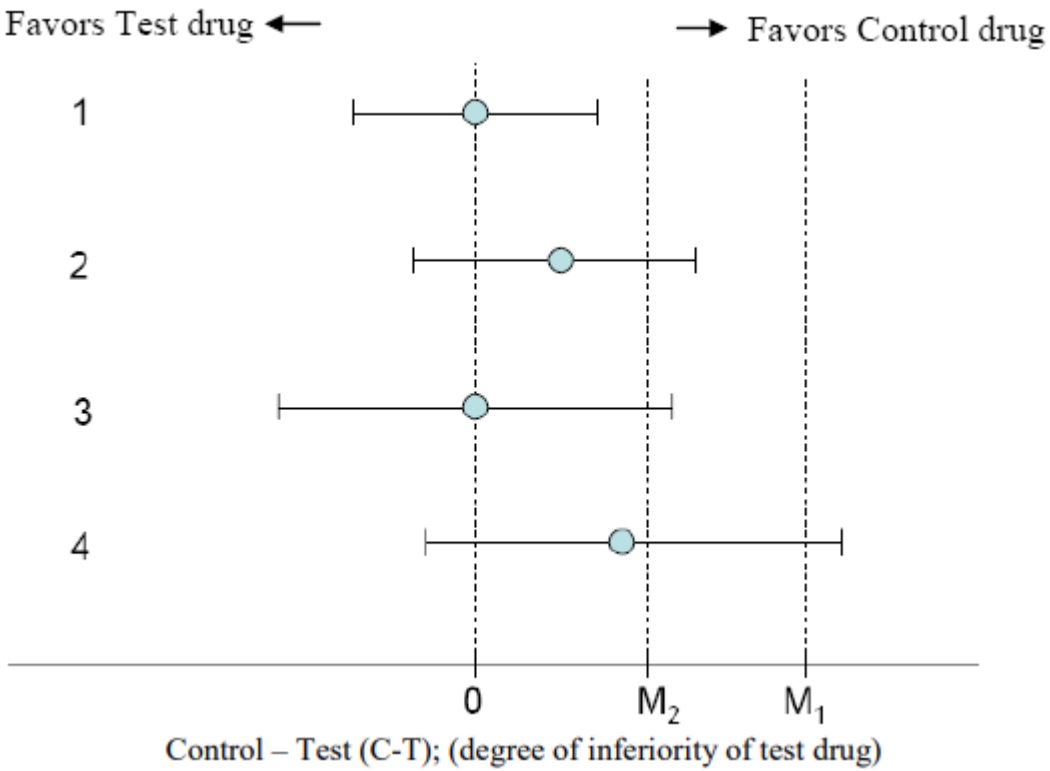
FDA guidance

- M_1 = the entire effect of the active control assumed to be present in the NI study
- M_2 = the largest clinically acceptable difference (degree of inferiority) of the test drug compared to the active control

- $M_2 \leq M_1$

- M_1 derived from historical trials
- M_2 clinical judgment

- Today we focus on M_1 only



Case study 1 – Weight management

- Imagine developing a drug in the same class as an already licensed weight management product. Similar efficacy is expected so a non-inferiority study is planned. How should the non-inferiority margin be chosen?
  - Patient population: obese or overweight with at least one risk factor
  - Endpoint: change from baseline to week 68
  - Summary measure: mean difference between active and control groups
  - Relevant intercurrent events: treatment discontinuation, rescue intervention (anti-obesity or bariatric surgery)

Case study 1 – Weight management

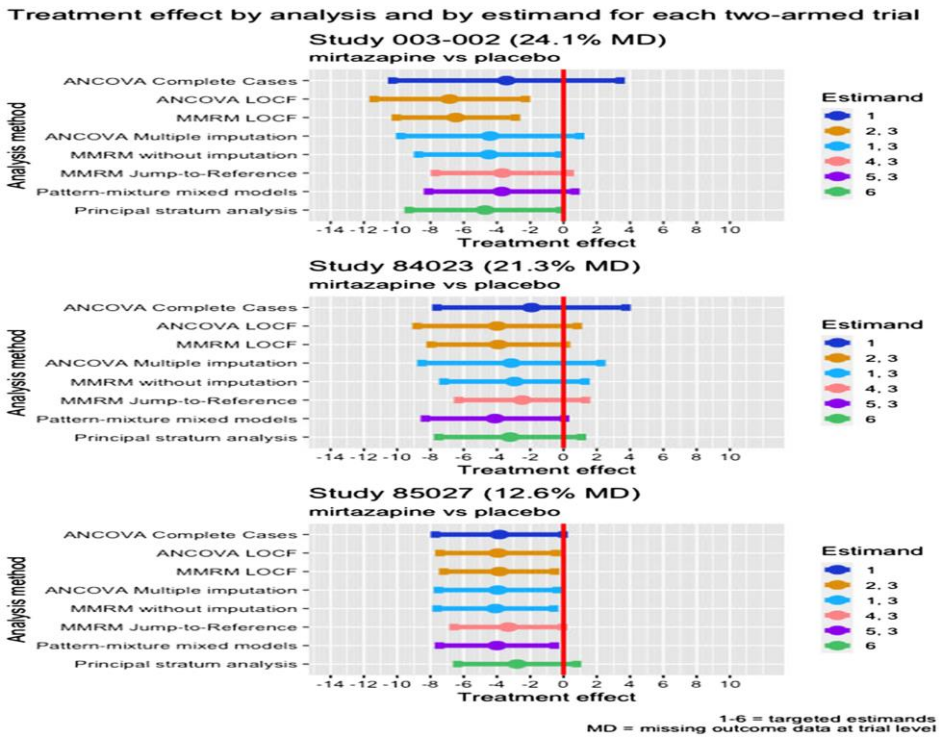
- Possible estimand strategies for 2 key intercurrent events were considered (treatment policy and hypothetical) [we could of course also choose treatment policy strategy for one and hypothetical for the other but this example is just for illustration]
- Different studies were of different durations
- Also some studies has slightly different patient populations
- Similar estimates obtained for treatment effect for treatment policy (11.4kg) and hypothetical estimands (12.1kg).
- So, in this case any disagreement on which estimand should be used is not crucial as they both yield very similar M1s and likely M2 will be much smaller than 11-12 kgs.

Case study 2 - Depression

- Estimation of treatment effects in short-term depression trials
- 6 RCTs that supported regulatory approval of mirtazapine. See paper by Mitroiu M, Teerenstra S, Oude Rengerink K, Pétavy F, Roes KCB. Estimation of treatment effects in short-term depression studies. An evaluation based on the ICH E9(R1) estimands framework. Pharmaceutical Statistics. 2022;21(5):1037–57 https://doi.org/10.1002/pst.2214

- Highlights the difficulty of understanding what the estimand for each trial might have been given it wasn't explained in the studies (as they were written before the estimand framework was in place).

Estimation of treatment effects in short-term depression studies. An evaluation based on the ICH E9(R1) estimands framework



Pharmaceutical Statistics, Volume: 21, Issue: 5, Pages: 1037-1057, First published: 09 June 2022, DOI: (10.1002/pst.2214)

Case study 2 - Depression

- More uncertainty on the estimand for each study.
- Making some assumptions difference between mirtazapine and placebo in change from baseline in the MADRS10 (Montgomery and Asberg Depression Rating Scale) between 3.6 and 3.23 depending on which estimand or set of different estimands used.
- More difficulty here knowing which estimand was targeted by each study. Able to do additional analyses in this case – this will most likely not be possible for other studies/medicinal products.

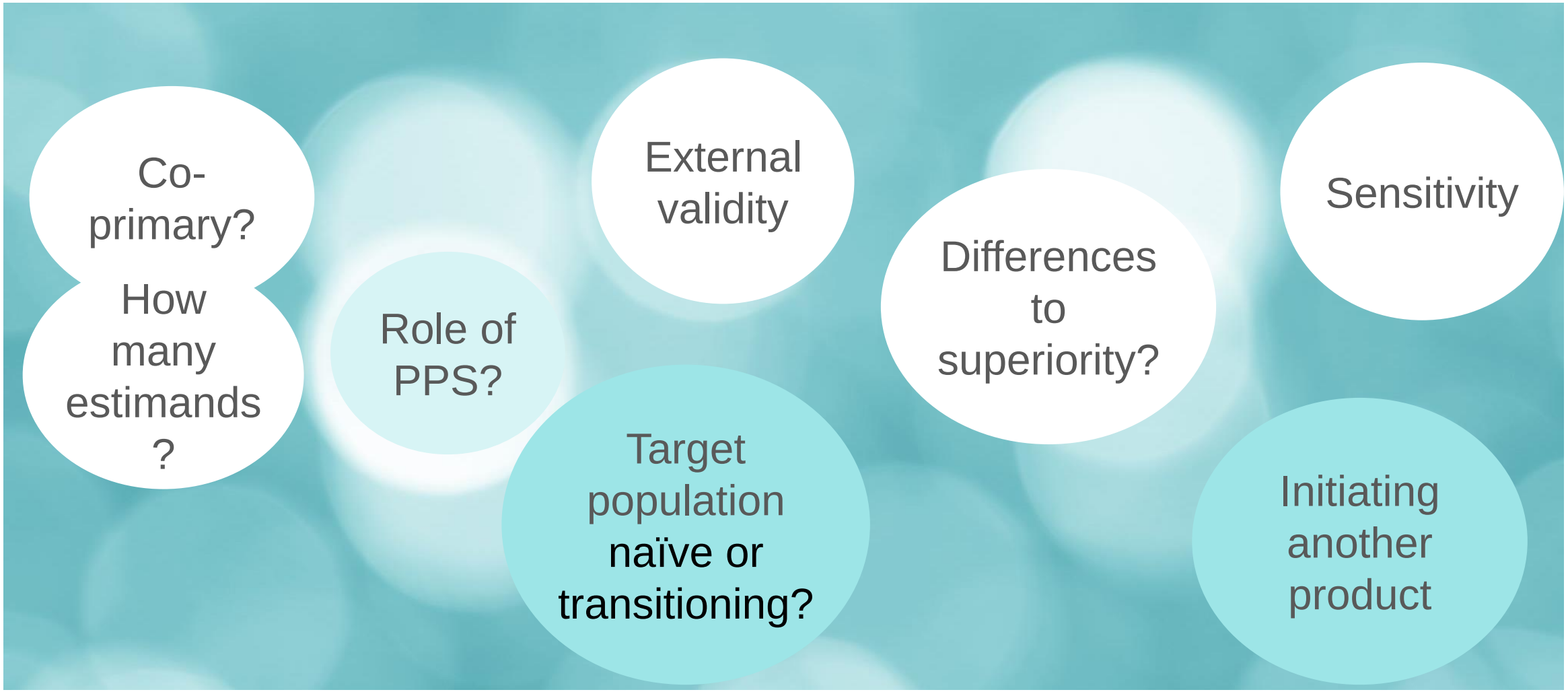
Key issues

- **Lack of information (often about intercurrent events) leading to uncertainty on what estimand was of interest in a study (pre E9 (R1) studies could not use E9 (R1))**
- **Different large studies having different estimands**
- **If individual patient level data (IPD) are not available to you - you can't conduct any new analyses**
  - Are simulations sufficient?
- **If IPD are available?**
- **All this work is only to provide indirect evidence that new drug is superior to placebo – so it shouldn't be a difficult exercise. Have to be careful we are not creating a lot of work when rough calculations may suffice.**

References

- Committee for Medicinal Products for Human use (CHMP). Guideline on the choice of non-inferiority margin. Published 2000. Accessed March 11, 2024. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf
- Food and Drug Administration (FDA). Non-inferiority clinical trials to establish effectiveness. Published 2016. https://www.fda.gov/media/78504/download
- Committee for Medicinal Products for Human Use (CHMP). Concept paper for the development of a guideline on non-inferiority and equivalence comparisons in clinical trials. Published 2024. https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-development-guideline-non-inferiority-equivalence-comparisons-clinical-trials_en.pdf
- Mitroiu M, Teerenstra S, Oude Rengerink K, Pétavy F, Roes KCB. Estimation of treatment effects in short-term depression studies. An evaluation based on the ICH E9(R1) estimands framework. Pharmaceutical Statistics. 2022;21(5):1037–57 https://doi.org/10.1002/pst.2214

Discussion topics



Discussion Questions

1. Role of PPS analysis in non-inferiority study.
    - What do you think is the role of PPS analysis in non-inferiority studies?

Discussion Questions

2. Multiple primary estimands

- In the past, non-inferiority was shown when the FAS and PPS analysis were statistically significant. Does this mean that in the future multiple primary estimands have to be specified? Are there particular settings/objectives in which you see value of multiple primary estimands? What are the settings in which a single primary estimand is sufficient?

Discussion Questions

3. “Sensitivity to detect differences”

- When defining an estimand for a non-inferiority trial, which roles does the concept of "sensitivity to detect differences" have as compared to the clinical relevance of an estimand? For example, choosing a treatment policy strategy for a particular intercurrent event might reflect clinical practice, but another strategy (e.g., composite strategy) might result in an estimand that is sensitive to detect differences.

Discussion Questions

4. Switching between non-inferiority and superiority

   - In studies which switch between non-inferiority and superiority, do you see the same estimand being used? Under what circumstances should they not be the same?

Discussion Questions

5. Non-inferiority margin

- It is recognized that the relevant non-inferiority margin depends on the estimand. At this time, many historical trials that might be used for calculating the margin are not using the estimand framework. How should sponsors reflect the estimand during the margin calculation when historical trials have not / only partially used the estimand framework?
- Similarly, if historical superiority trials address a treatment policy estimand, as is the case in many superiority trials, but the estimand in the non-inferiority trial will not be addressing a treatment policy estimand, how should this difference in estimands be reflected in the calculation of the non-inferiority margin?

Potential Discussion Questions

1. Role of PPS analysis in non-inferiority study.
    - What do you think is the role of PPS analysis in non-inferiority studies?
2. Multiple primary estimands
    - In the past, non-inferiority was shown when the FAS and PPS analysis were statistically significance. Does this mean that in future multiple primary estimands have to be specified? Are there particular settings/objectives in which you see value of multiple primary estimands? What are the settings in which a single primary estimand is sufficient?
3. “Sensitivity to detect differences”
    - When defining an estimand for a non-inferiority trial, which roles does the concept of "sensitivity to detect differences" have as compared to the clinical relevance of an estimand? For example, choosing a treatment policy strategy for a particular intercurrent event might reflect clinical practice, but another strategy (e.g., composite strategy) might be result in an estimand that is sensitive to detect differences.
4. Switching between non-inferiority and superiority
    - In studies which switch between non-inferiority and superiority, do you see the same estimand being used? Under what circumstances should they not be the same?
5. Non-inferiority margin
    - It is recognized that the relevant non-inferiority margin depends on the estimand. At this time, many historical trials that might be used for calculating the margin are not using the estimand framework. How should sponsors reflect the estimand during the margin calculation when historical trials have not / only partially used the estimand framework?
    - Similarly, if historical superiority trials address a treatment policy estimand, as is the case in many superiority trials, but the estimand in the non-inferiority trial will not be addressing a treatment policy estimand, how should this difference in estimands be reflected in the calculation of the non-inferiority margin?