Determining the non-inferiority margin in light of the estimand framework

Sunita Rehal, GSK;

Helle Lynggaard, Novo Nordisk; Marian Mitroiu, Biogen; Tobias Mütze, Novartis;
David Wright, AstraZeneca

11 June 2025 PSI Annual Conference, London, UK

Estimands in non-inferiority trials
Disclaimer

Disclaimer:

The views expressed by the presenter and co-authors are not necessarily the views and practices of their employers, or of any of the EIWG member companies and organisations.

Acknowledgements

- Chein-Ju Lin (Roche)
- Claudia Hemmelmann (Sandoz)
- Oliver Keene (KeeneOnStatistics), and
- EIWG sub-team on estimands in non-inferiority trials: EFSPI/EFPIA EIWG - Sub-teams

Following on from our previous paper

MAIN PAPER OPEN ACCESS

Applying the Estimand Framework to Non-Inferiority Trials

Helle Lynggaard[1] | Oliver N. Keene[2] | Tobias Mütze[3] | Sunita Rehal[4]

[1]Novo Nordisk A/S, Bagsvaerd, Denmark | [2]KeeneONStatistics, Maidenhead, UK | [3]Novartis Pharma AG, Basel, Switzerland | [4]GlaxoSmithKline plc, Brentford, UK

Correspondence: Helle Lynggaard (hlyn@novonordisk.com)

Received: 14 December 2023 | **Revised:** 15 July 2024 | **Accepted:** 24 July 2024

Funding: GSK and Novo Nordisk each sponsored half of the open access fees.

Keywords: clinical question of interest | estimand framework | ICH E9(R1) | non-inferiority | regulatory guidance | superiority

ABSTRACT

Most published applications of the estimand framework have focused on superiority trials. However, non-inferiority trials present specific challenges compared to superiority trials. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use notes in their addendum on estimands and sensitivity analysis in clinical trials that there may be special considerations to the implementation of estimands in clinical trials with a non-inferiority objective yet provides little guidance. This paper discusses considerations that trial teams should make when defining estimands for a clinical trial with a non-inferiority objective. We discuss how the pre-addendum way of establishing non-inferiority can be embraced by the estimand framework including a discussion of the role of the Per Protocol analysis set. We examine what clinical questions of interest can be formulated in the context of non-inferiority trials and outline why we do not think it is sensible to describe an estimand as 'conservative'. The impact of the estimand framework on key considerations in non-inferiority trials such as whether trials should have more than one primary estimand, the choice of non-inferiority margin, assay sensitivity, switching from non-inferiority to superiority and estimation are discussed. We conclude by providing a list of recommendations, and important considerations for defining estimands for trials with a non-inferiority objective.

Considering estimands for evidence synthesis in non-inferiority studies

- *"For **synthesising evidence across clinical trials**, **the same estimand should be considered** at the planning stage of the contributing trials. Similar considerations apply, for example, to the design of a meta-analysis, **using estimated effect sizes from completed trials to determine non-inferiority margins,** or the use of external control groups for the interpretation of single-arm trials."* - ICH E9(R1) training slides.

Considering estimands for evidence synthesis in non-inferiority studies

- *"For **synthesising evidence across clinical trials**, **the same estimand should be considered** at the planning stage of the contributing trials. Similar considerations apply, for example, to the design of a meta-analysis, **using estimated effect sizes from completed trials to determine non-inferiority margins,** or the use of external control groups for the interpretation of single-arm trials."* - ICH E9(R1) training slides.

- Guidance from the EMA on non-inferiority and equivalence will be updated:
  - Concept paper for the development of a guideline on non-inferiority and equivalence comparisons in clinical trials
    - The Guideline on the Choice of Non-Inferiority Margin, and the Points to consider on Switching between Superiority and Non-Inferiority, will be merged into one.

Considering estimands for evidence synthesis in non-inferiority studies

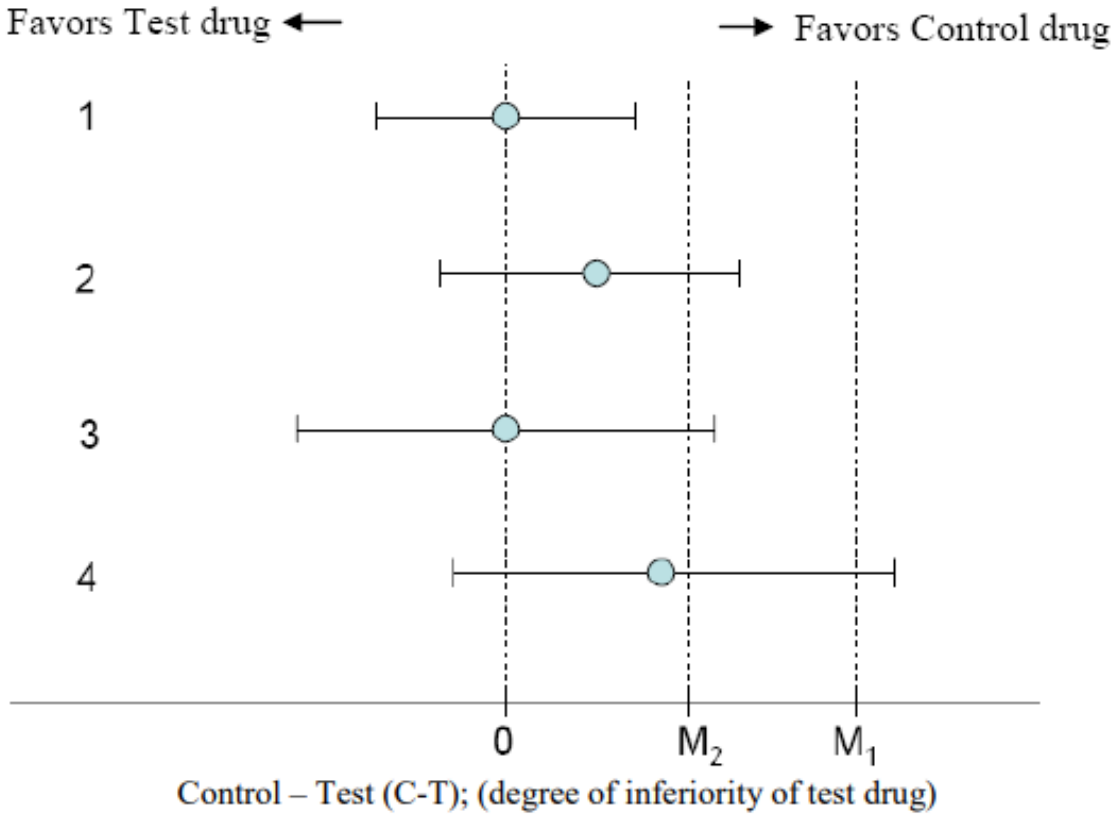
- *"For **synthesising evidence across clinical trials**, **the same estimand should be considered** at the planning stage of the contributing trials. Similar considerations apply, for example, to the design of a meta-analysis, **using estimated effect sizes from completed trials to determine non-inferiority margins,** or the use of external control groups for the interpretation of single-arm trials."* - ICH E9(R1) training slides.

- Guidance from the EMA on non-inferiority and equivalence will be updated:
  - Concept paper for the development of a guideline on non-inferiority and equivalence comparisons in clinical trials
    - The Guideline on the Choice of Non-Inferiority Margin, and the Points to consider on Switching between Superiority and Non-Inferiority, will be merged into one.

- The margin defined will be specific to the estimand:
  - The margin is used as a threshold for comparing the effect being estimated.
  - The margin should be aligned with the population and relevant intercurrent events.

How the margin could be calculated (FDA guidance)

- M_1 = the entire effect of the active control (reference) assumed to be present in the non-inferiority 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_2 more stringent in this example.
- **M_1 derived from historical trials** (statistical judgment).
- M_2 clinical judgment.
- Meta-analyses usually conducted.
  - Based on superiority testing here.
- New study will test for non-inferiority.
- It is likely information used will come from different estimands.



When the exact estimand is unknown from previous studies

- Unclear how to include information from historical studies.
  - Studies of interest often performed pre-addendum.
  - Estimand not explicitly defined leading to reference studies potentially with different estimands combined.
    - The impact of this is likely to depend on what those differences are.

- In some cases, we might be able to exploit the information available:
  - CONSORT flow charts/the statistical analysis within publications to help derive the underlying estimand.

Example – Determining M1 for a new study (weight management)

- A new study is being considered for weight management where the new treatment being developed is in the same class as a licensed product currently on market and expected to have a similar efficacy.
- Two intercurrent events: discontinuation and rescue therapy.
- Using a treatment policy strategy to handle intercurrent events may not be the most relevant for a non-inferiority study.

Example – Determining M1 for a new study (weight management)

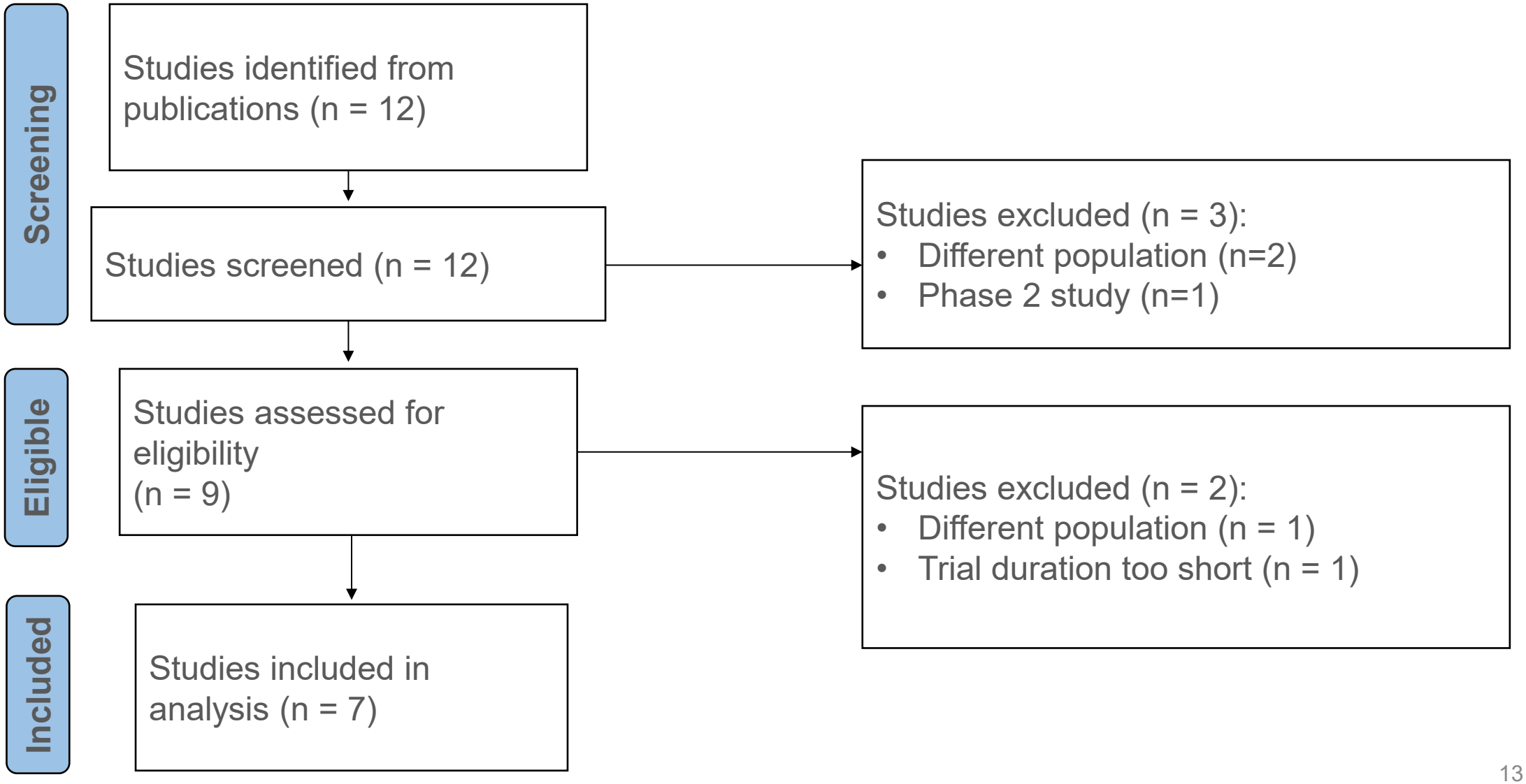
- A new study is being considered for weight management where the new treatment being developed is in the same class as a licensed product currently on market and expected to have a similar efficacy.
- Two intercurrent events: discontinuation and rescue therapy.
- Using a treatment policy strategy to handle intercurrent events may not be the most relevant for a non-inferiority study.
- For weight management, use of rescue may include additional interventions that help weight loss.
  - E.g. major changes to exercise patterns or diet, bariatric surgery.
  - Using a treatment policy strategy could understate (or overstate) the benefit (or harm) of the new treatment.
- It may be useful to target the effect including any subsequent effects of discontinuation and assuming no rescue therapy had been applied.

Example – Primary estimand for a new study (weight management)

Primary estimand:

Attributes	Primary
Population	Adult patients with a BMI ≥ 30 (or ≥ 27 if patients had ≥1 weight-related co-existing condition) who do not have diabetes.
Endpoint	Relative change from baseline to week 68 in body weight (%).
Treatment	New treatment vs. once-weekly subcutaneous semaglutide (2.4 mg), including any subsequent effects of discontinuing treatment and assuming rescue medication did not occur.
Population-level summary	Difference in mean.
Intercurrent events (handling strategy)	• Treatment discontinuation: treatment policy. • Rescue therapy: hypothetical.

Studies included in the meta-analysis

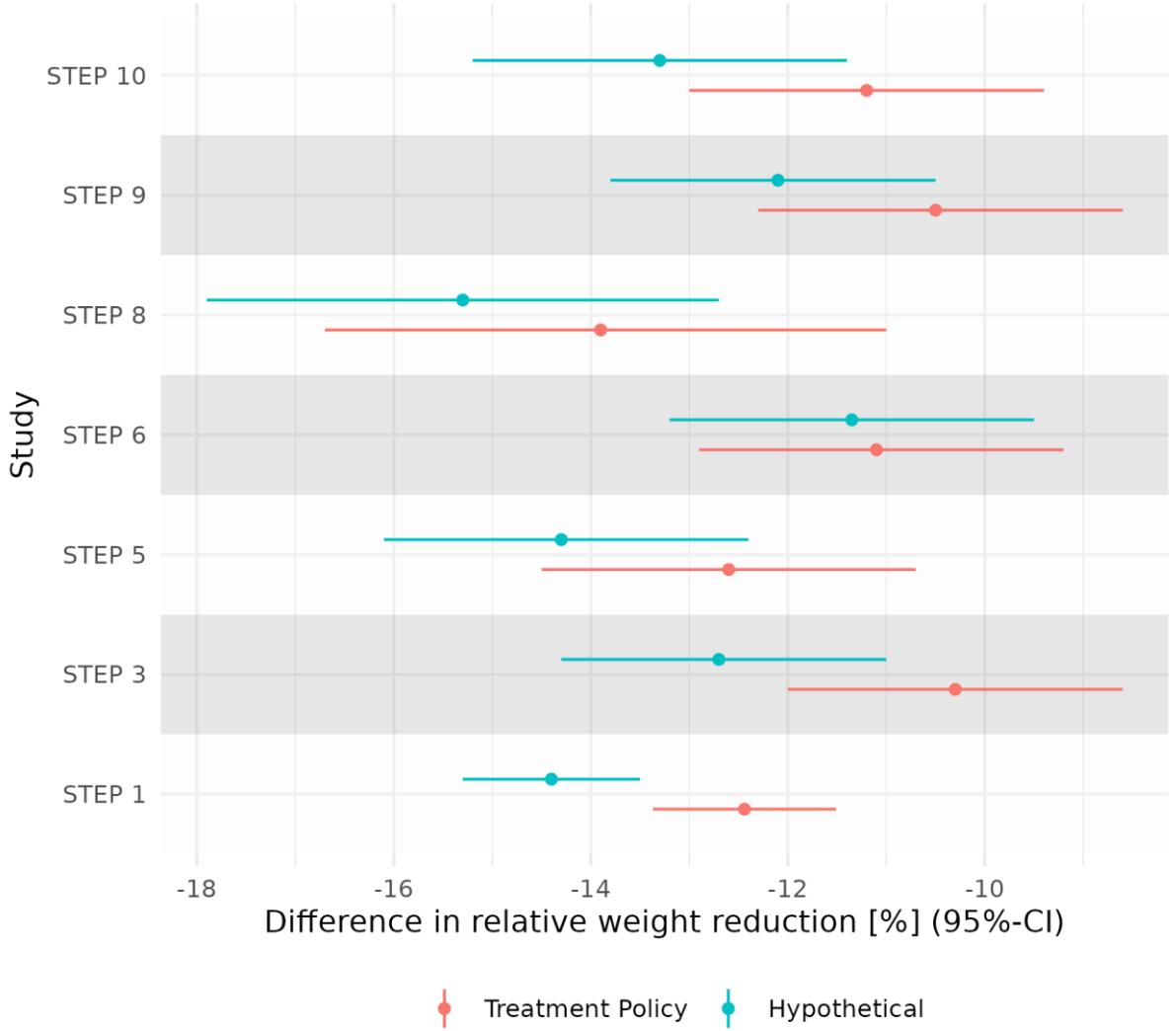


Example - Weight management (estimand from historical studies)

- Based on STEP studies.
- Two intercurrent events: treatment discontinuation and rescue therapy.
- Not exactly aligned with the estimand specified for our new study.
- Primary and supplementary estimands:

Attributes	Primary	Supplementary
Population	Adult patients with a BMI ≥ 30 (or ≥ 27 if patients had ≥1 weight-related co-existing condition) who do not have diabetes.	
Endpoint	Relative change from baseline to week 68 in body weight (%).	
Treatment	Once-weekly subcutaneous semaglutide (2.4 mg) vs. placebo.	
Population-level summary	Difference in mean.	
Intercurrent events (handling strategy)	• Treatment discontinuation: Treatment policy. • Rescue therapy: Treatment policy.	• Treatment discontinuation: hypothetical. • Rescue therapy: hypothetical.

Placebo-adjusted treatment effects from STEP trials



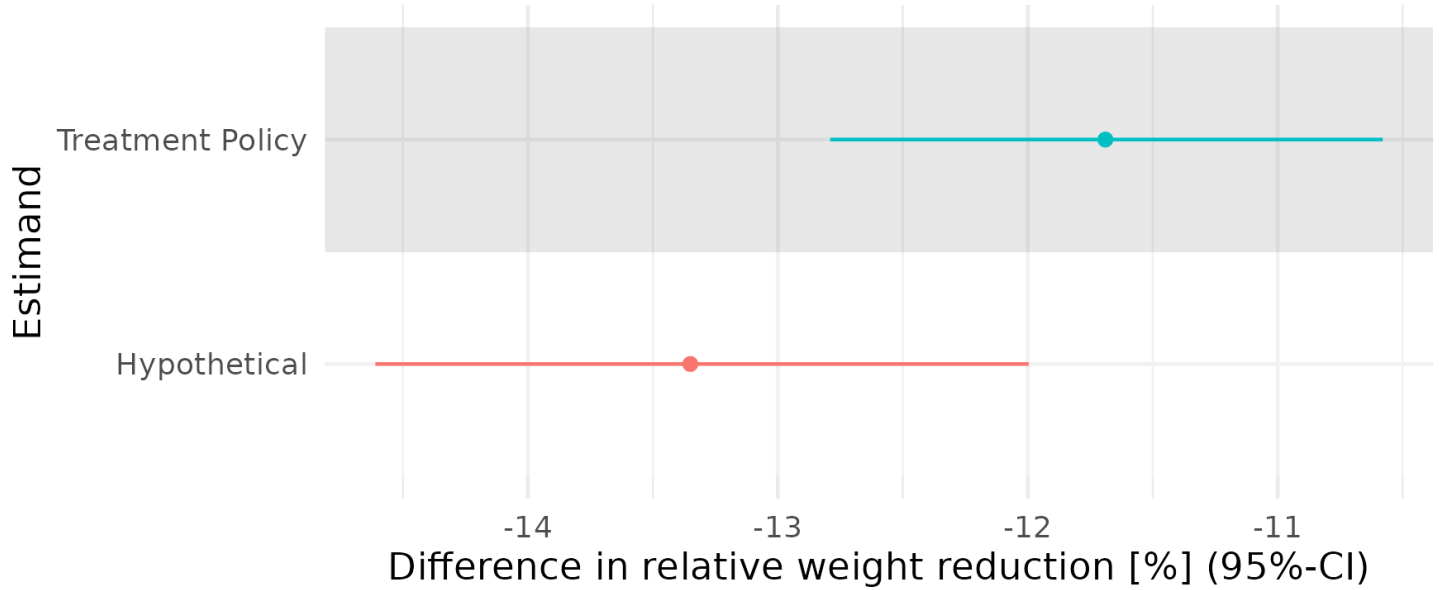
Available information from patient flow-chart

- Can we exploit any additional information from the CONSORT flow chart to help?

Study	Intercurrent event			
	Treatment discontinuation		Rescue therapy	
	Semaglutide (n/N)	Placebo (n/N)	Semaglutide (n/N)	Placebo (n/N)
STEP 1	223/1306	147/655	7/1306	13/655
STEP 3	31/407	13/204	.	.
STEP 6	13/199	3/101	.	.
STEP 8	18/126	15/85	1/126	3/85
STEP 9	34/271	29/136	.	.
STEP 10	18/138	8/69	.	.

- Even when estimands are fully defined, the full picture may still be unknown.

Results of meta-analysis from STEP studies



Handling strategy for discontinuation and rescue therapy	**Pooled % mean treatment-effect**
Treatment policy	-11.7 (-12.7, -10.5)
Hypothetical	-13.4 (-14.6, -12.0)

What have we learned from this exercise?

- The choice of the margin is arbitrary where the estimand to establish M1 is unclear.

What have we learned from this exercise?

- The choice of the margin is arbitrary where the estimand to establish M1 is unclear.
- To provide sufficient evidence of superiority of the active control arm over placebo:
  - Exploit the use of CONSORT flow chart in published trials (identification of intercurrent events).
  - Discussion of what intercurrent events are relevant within study teams (proportion of expected intercurrent events).

What have we learned from this exercise?

- The choice of the margin is arbitrary where the estimand to establish M1 is unclear.
- To provide sufficient evidence of superiority of the active control arm over placebo:
  - Exploit the use of CONSORT flow chart in published trials (identification of intercurrent events).
  - Discussion of what intercurrent events are relevant within study teams (proportion of expected intercurrent events).
- Check the sensitivity of the choice of margin.
  - What the impact on the margin is under different assumptions.

What have we learned from this exercise?

- The choice of the margin is arbitrary where the estimand to establish M1 is unclear.
- To provide sufficient evidence of superiority of the active control arm over placebo:
  - Exploit the use of CONSORT flow chart in published trials (identification of intercurrent events).
  - Discussion of what intercurrent events are relevant within study teams (proportion of expected intercurrent events).
- Check the sensitivity of the choice of margin.
  - What the impact on the margin is under different assumptions.
- Important to consider whether estimands from historical studies are consistent with the estimand for a new study.
  - Including whether the same estimand could be used for superiority after demonstrating non-inferiority.

What have we learned from this exercise?

- The choice of the margin is arbitrary where the estimand to establish M1 is unclear.
- To provide sufficient evidence of superiority of the active control arm over placebo:
  - Exploit the use of CONSORT flow chart in published trials (identification of intercurrent events).
  - Discussion of what intercurrent events are relevant within study teams (proportion of expected intercurrent events).
- Check the sensitivity of the choice of margin.
  - What the impact on the margin is under different assumptions.
- Important to consider whether estimands from historical studies are consistent with the estimand for a new study.
  - Including whether the same estimand could be used for superiority after demonstrating non-inferiority.
- Before finalising the choice of the proposed margin, clinical judgement (M2) with respect to the estimand is also required.
  - Is the margin sensible for the study and is it clinically meaningful?

Ongoing/future work

- M1 is determined from previous superiority placebo-controlled trials and used as a benchmark comparing it with an effect that targets a different clinical question for a different hypothesis.
  - How can this be meaningfully combined?

Ongoing/future work

- M1 is determined from previous superiority placebo-controlled trials and used as a benchmark comparing it with an effect that targets a different clinical question for a different hypothesis.
  - How can this be meaningfully combined?

Investigating whether simulation based on historical information on intercurrent events from superiority studies would be a good approach:

- Can we use information on the distribution of intercurrent events in combination with clinical judgment?
  - There might be strong assumptions attached as there are lots of moving parts to consider.
    - Proportion of intercurrent events.
    - Frequency of intercurrent events over time.
    - The meaning of intercurrent events could change over time, e.g. rescue medication/therapy.

Ongoing/future work

- M1 is determined from previous superiority placebo-controlled trials and used as a benchmark comparing it with an effect that targets a different clinical question for a different hypothesis.
  - How can this be meaningfully combined?

Investigating whether simulation based on historical information on intercurrent events from superiority studies would be a good approach:

- Can we use information on the distribution of intercurrent events in combination with clinical judgment?
  - There might be strong assumptions attached as there are lots of moving parts to consider.
    - Proportion of intercurrent events.
    - Frequency of intercurrent events over time.
    - The meaning of intercurrent events could change over time, e.g. rescue medication/therapy.

- Some of these issues will be addressed within a paper.

Final thoughts

- When the margin cannot be determined, maintain best efforts to utilise any relevant data available for the reference arm as far as possible.

Final thoughts

- When the margin cannot be determined, maintain best efforts to utilise any relevant data available for the reference arm as far as possible.
- Additional assumptions will need to be made when relevant data from historical studies is unavailable.
  - Initiate study team discussions on the choice of M1.

Final thoughts

- When the margin cannot be determined, maintain best efforts to utilise any relevant data available for the reference arm as far as possible.
- Additional assumptions will need to be made when relevant data from historical studies is unavailable.
  - Initiate study team discussions on the choice of M1.
- Important to be transparent about the assumptions made.
  - Any limitations of determining the margin due to misalignment with the estimand should be acknowledged.
  - Acknowledge whether handling of intercurrent events differ or whether slightly different intercurrent events were identified in previous studies.

Final thoughts

- When the margin cannot be determined, maintain best efforts to utilise any relevant data available for the reference arm as far as possible.
- Additional assumptions will need to be made when relevant data from historical studies is unavailable.
  - Initiate study team discussions on the choice of M1.
- Important to be transparent about the assumptions made.
  - Any limitations of determining the margin due to misalignment with the estimand should be acknowledged.
  - Acknowledge whether handling of intercurrent events differ or whether slightly different intercurrent events were identified in previous studies.
- Important to get agreement on the choice of the margin between the sponsor and regulator prior to conducting the study.
  - This is critical when there is limited historical information available.

Final thoughts

- When the margin cannot be determined, maintain best efforts to utilise any relevant data available for the reference arm as far as possible.
- Additional assumptions will need to be made when relevant data from historical studies is unavailable.
  - Initiate study team discussions on the choice of M1.
- Important to be transparent about the assumptions made.
  - Any limitations of determining the margin due to misalignment with the estimand should be acknowledged.
  - Acknowledge whether handling of intercurrent events differ or whether slightly different intercurrent events were identified in previous studies.
- Important to get agreement on the choice of the margin between the sponsor and regulator prior to conducting the study.
  - This is critical when there is limited historical information available.
- Part of the EMA 3-year methodology working plan (2025-2027) is to provide guidance on aligning estimand attributes across different studies in the context of meta-analysis.
  - Methodology Working Party Workplan 2025-2027 – Adopted.

Questions