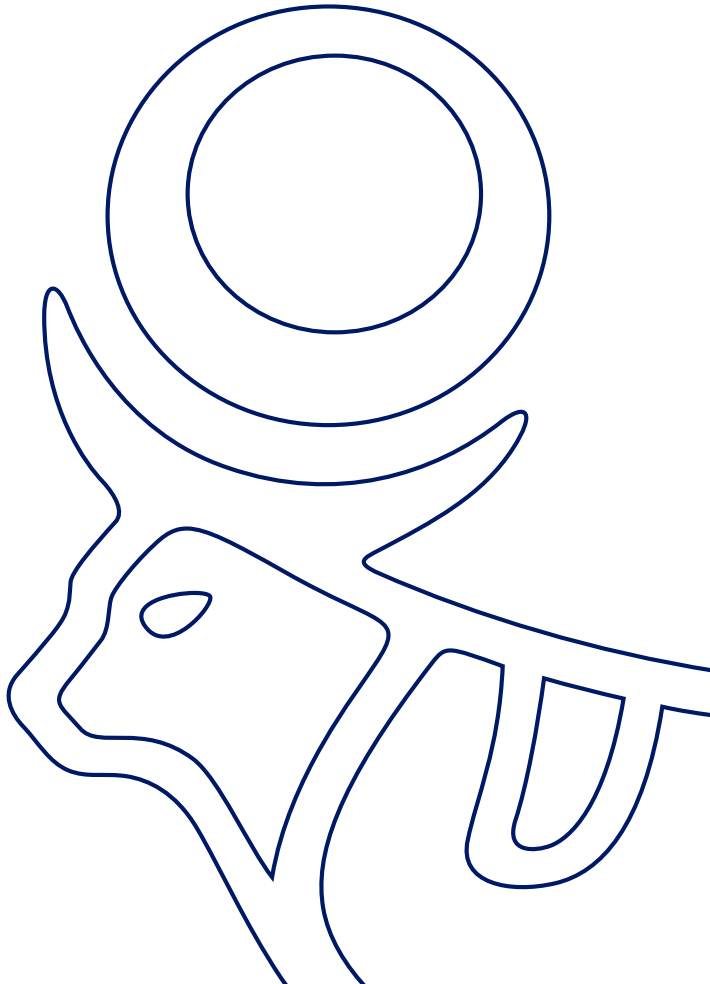
UNORDINARY DRIVES CHANGE

novo nordisk®

Evidence synthesis

Estimands, transportability and external validity



Antonio Remiro -Azócar, PhD

Methods and Outreach, Novo Nordisk

10th EFSPI Regulatory Statistics Workshop

10th September 2025

Disclaimers

This presentation is based on the following articles:

- Remiro-Azócar, A. et al (EFPIA/EFSPI Estimands Implementation Working Group sub-team on estimands in late phase), 2025. Incorporating estimands into meta-analyses of clinical trials. Pre-print TBA soon.
- Phillippo, D.M., Remiro-Azócar, A., Heath, A., Baio, G., Dias, S., Ades, A.E. and Welton, N.J., 2025. Effect modification and non-collapsibility together may lead to conflicting treatment decisions: A review of marginal and conditional estimands and recommendations for decision-making. Research Synthesis Methods, 16(2), pp.323-349.
- Remiro-Azócar, A., 2024. Transportability of model-based estimands in evidence synthesis. Statistics in Medicine, 43(22), pp.4217-4249.
- Remiro-Azócar, A., Phillippo, D.M., Welton, N.J., Dias, S., Ades, A.E., Heath, A. and Baio, G., 2025. Marginal and conditional summary measures: transportability and compatibility across studies. arXiv preprint arXiv:2507.21925.
- Remiro-Azócar, A., 2022. Target estimands for population-adjusted indirect comparisons. Statistics in Medicine, 41(28), pp.5558-5569.
- Remiro-Azócar, A. and Gorst-Rasmussen, A., 2024. Broad versus narrow research questions in evidence synthesis: a parallel to (and plea for) estimands. Research Synthesis Methods, 15(5), pp.735-740.

The views and opinions expressed herein are solely those of the presenter and are not necessarily those of Novo Nordisk A/S. Any of these views and opinions cannot and should not necessarily be construed to represent those of Novo Nordisk A/S or its affiliates.

Acknowledgements

Estimands Implementation Working Group sub -team on estimands in late phase

Alessandro Ghiretti Arthur Allignol	Daiichi-Sankyo
Amel Besseghir	Independent Consultant
Khadija Rantell	Medicines & Healthcare products Regulatory Agency
Frank Kleinjung	
Lotte Husemoen Pepa Polavieja Antonio Remiro-Azócar	novo nordisk®
Antonia Morga	astellas
Barbara Rosettani	Independent Consultant

Tatsiana Vaitsiakhovich	
Emmanuelle Boutmy	Merck KGaA, Darmstadt, Germany

Other co-authors:

A. E. Ades, Anders Gorst-Rasmussen, Anna Heath, David M. Phillippo, Gianluca Baio, Helle Lynggaard, Jay J. H. Park, Nicky J. Welton, Sofia Dias

Other acknowledgements:

Novo Nordisk Evidence Synthesis Scientific Working Group and Estimands Scientific Working Group

Background

High-level summary

In health technology assessment (HTA), evidence synthesis methods inform policy and reimbursement decision-making, for specific clinical settings with well-defined research questions

Base-case scenario:

- A pairwise or network meta-analysis, combining several RCTs, is required for HTA
- Each RCT has been designed to support regulatory approval and has target estimands of its own

Two major challenges:

(1) Heterogeneity or **lack of transportability** between trials addressing different research questions

(2) Ambiguous **external validity** of pooled estimates relative to the decision-making context

Question:

- Can the estimand framework – ICH E9 (R1) definition – be used to resolve these challenges?

Not just a HTA issue

ICH E9 (R1) addendum:

- Lacks explicit guidelines for estimands in meta-analysis
- But already warns against a "naïve comparison of data sources, or integration of data from multiple trials without consideration of the estimand that is addressed in each"

Draft consolidated 3-year rolling work plan (2026-28) from the EMA Methodology Working Party:

- Lists "Guidance on how to align estimand attributes across different trials in the context of a meta-analysis" as an activity to be started

PICO – a broad research question

PICO extracted from the final scope of the NICE health technology appraisal of cabozantinib:

Population	Adults with locally advanced or metastatic differentiated thyroid carcinoma, whose disease is refractory to, or who are unsuitable for radioactive iodine, and whose disease has progressed during or after prior systemic therapy
Intervention	Cabozantinib
Comparator	Best supportive care
Outcomes	Overall survival Progression-free survival Response rate Adverse effects of treatment Health-related quality of life

https://www.nice.org.uk/guidance/ta928/resources/cabozantinib-for-previously-treated-advanced-differentiated-thyroid-cancer-unsuitable-for-or-refractory-to-radioactive-iodine-pdf-82615554322117

- Population is narrowly defined, based on the claimed therapeutic indication
- Intervention/comparator are broad: dosage, regimen or mode of administration not specified
- Outcomes are broad: outcome measure instruments and time points not specified
- Does not consider population-level summary measure or intercurrent event strategies as part of the scope

Estimands – a more precise research question

Primary estimands for the COSMIC-311 Phase 3 RCT, used to obtain marketing authorization for cabozantinib:

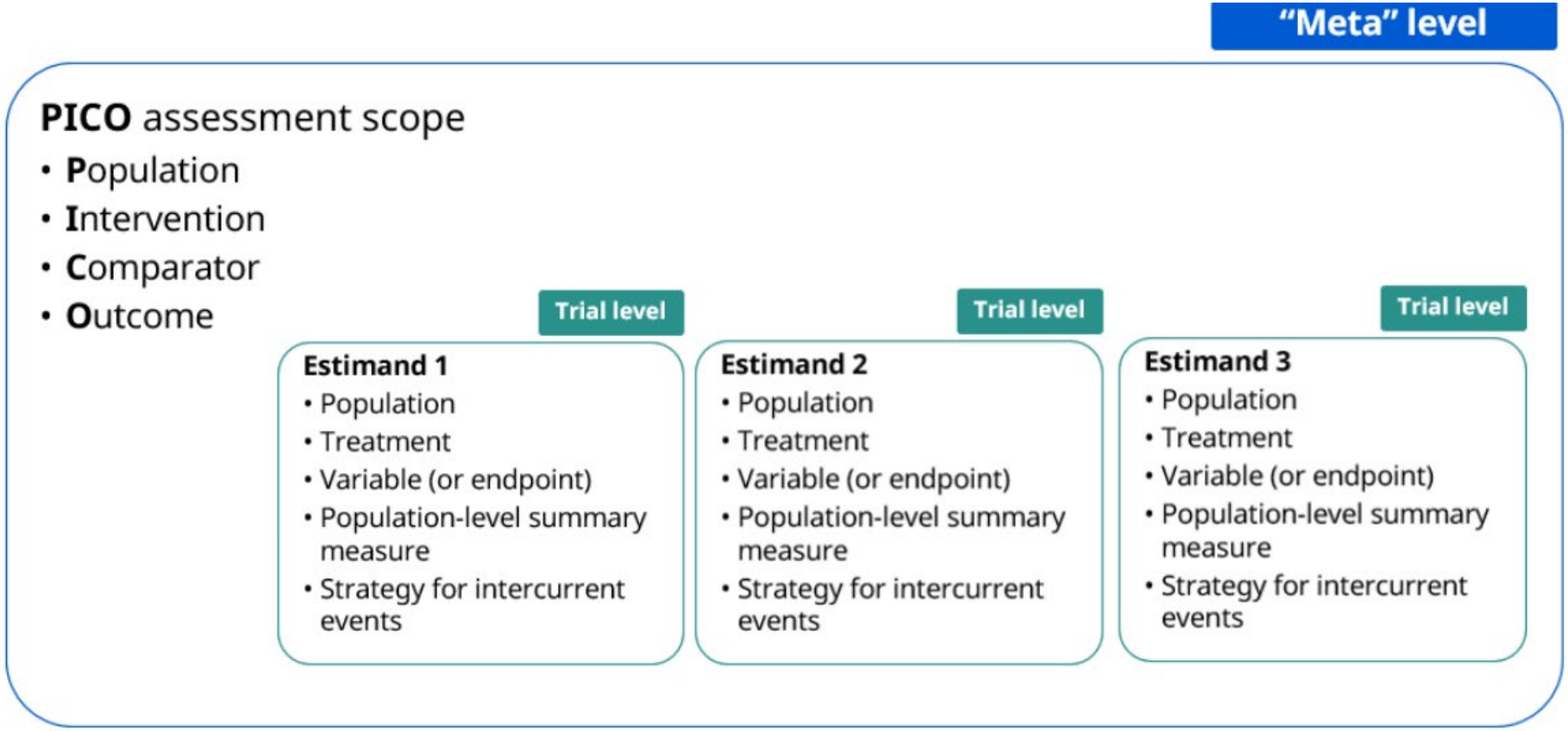
- Population narrowly defined
- Treatments narrowly defined: include dosing, regimen, and mode of administration
- Variable/endpoint includes specific outcome measures or instruments
- Population-level summary measure is specified
- Intercurrent event strategies are explicitly addressed

Estimand	Objective response rate	Progression-free survival
Population	Patients with radioiodine-refractory differentiated thyroid cancer who have progressed after prior VEGFR-targeted therapy	
Treatments	Oral cabozantinib (60 mg once daily) Matching placebo	
Variable (endpoint)	Radiographic response per RECIST 1.1	Duration of radiographic progression-free survival
Population-level summary measure	Difference in proportions of subjects with a best overall response of confirmed complete response or confirmed partial response per RECIST 1.1 between treatment conditions	Difference in survival functions between treatment conditions
Intercurrent event strategies	**Treatment policy** for receipt of local radiation to bone, surgical resection of non-target tumor lesions, death, loss to radiographic follow-up, or receipt of local non-protocol anti-cancer medications other than for disease under study **While on treatment** for surgical resection of target tumor lesions, receipt of systemic non-protocol anti-cancer medications, local non-protocol anti-cancer medications for disease under study, or local radiation to soft tissue for disease under study	**Treatment policy** for clinical deterioration, receipt of local radiation to bone, surgical resection of non-target tumor lesions, or receipt of local non-protocol anti-cancer medications other than for disease under study **Hypothetical** for surgical resection of target tumor lesions, receipt of systemic non-protocol anti-cancer medications, local non-protocol anti-cancer medications for disease under study, or local radiation to soft tissue for disease under study

https://www.ema.europa.eu/en/documents/variation-report/cabometyx-h-c-004163-ii-0023-epar-assessment-report-variation_en.pdf

Estimands as complements to PICOs

- Estimand attributes are typically encompassed by their corresponding PICO elements
- PICOs can be viewed as sets of different estimands, with a given PICO potentially containing multiple estimand definitions



Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials.

The case for broad research questions

Estimands – selected by sponsors – in agreement with regulators – that can align their trial's design, data collection and analysis with a desired research question

PICOs – policy-driven, not necessarily driven by the available data, and determined after some or most trials have been conducted

In the evidence synthesis context, a broad research question…

- Describes the "totality" of evidence for **qualitative** systematic literature reviews
- Enhances the feasibility of "anchoring" the evidence network for indirect treatment comparisons
- Facilitates the consolidation of HTA scopes, e.g., EU Joint Clinical Assessment

PICOs are used to pose research questions in evidence synthesis: to guide the data extraction process for systematic literature reviews and to determine the scope of health technology assessments

The case for narrow research questions

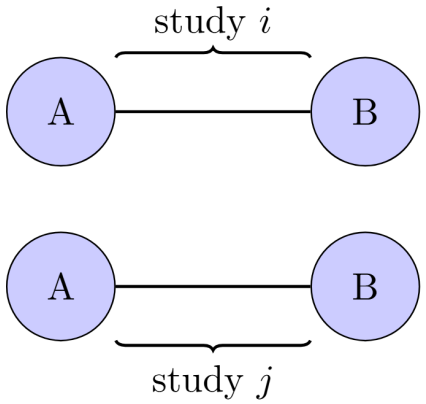
Healthcare decision-making requires taking a narrow perspective of research questions..to inform policy decisions for specific healthcare settings, treatment strategies, target populations and outcomes

Two major threats for the validity of **quantitative** evidence syntheses:

1. Heterogeneity or **lack of transportability** across trials addressing different research questions

Pairwise meta-analysis of direct comparisons

$$\Delta_{AB}^{(i)} \neq \Delta_{AB}^{(j)}$$

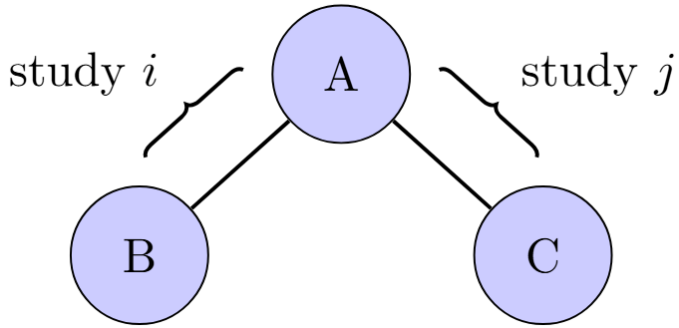


Network meta-analysis and indirect comparisons

$$\Delta_{AB}^{(i)} \neq \Delta_{AB}^{(j)}$$

or

$$\Delta_{AC}^{(i)} \neq \Delta_{AC}^{(j)}$$



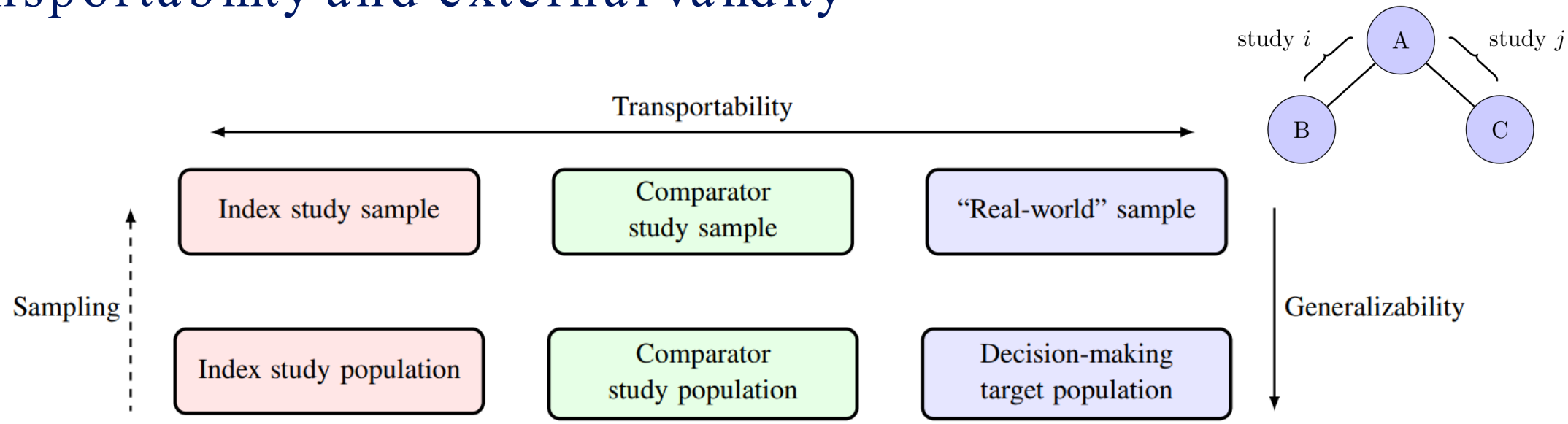
2. **Limited external validity** or applicability of pooled estimates

$$\frac{\sum_{i=1}^{K} \Delta_{AB}^{(i)} w_i}{w_i} \neq \Delta_{AB}^{(target)}$$

$$\frac{\sum_{i=1}^{K} \Delta_{AC}^{(i)} w_i}{w_i} \neq \Delta_{AC}^{(target)}$$

$$\frac{\sum_{i=1}^{K} \Delta_{BC}^{(i)} w_i}{w_i} \neq \Delta_{BC}^{(target)}$$

Transportability and external validity



Remiro-Azócar, A., 2022. Target estimands for population-adjusted indirect comparisons. Statistics in Medicine, 41(28), pp.5558-5569.

These terms normally refer to the "population" attribute of estimands

We take a more general view, incorporating estimand attributes not necessarily considered by PICO:

- **Population -level summary measure** (causal contrast)
- **Intercurrent event strategies** (treatment strategies)

Estimands for meta-analysis: a proposal

(1) Estimands **at the trial level** should be considered to:

- Identify, explain and potentially mitigate heterogeneity between trials
- Assess transportability across trials
- Make cross-trial misalignments explicit to avoid mixing "apples and oranges"

(2) Estimands **at the meta -analytical level** :

- Can potentially be defined "pragmatically"..possible divergence from the "estimand thinking process"
- Can optimize the external validity of pooled estimates relative to the decision-making context

The (population-level) summary measure (i.e., the causal contrast)

Trial-level estimands

Summary measures can be marginal or conditional, collapsible or non-collapsible, directly collapsible…

Marginal treatment effect

$$\text{MTE} = g\left(E\left(Y^1\right)\right) - g\left(E\left(Y^0\right)\right)$$

Population -average conditional treatment effect

$$\text{PACTE} = E_X\left(g\left(E\left(Y^1 \mid X = x\right)\right) - g\left(E\left(Y^0 \mid X = x\right)\right)\right)$$

Conditional average treatment effect *

$$\text{CATE} = g\left(E\left(Y^1 \mid X = x\right)\right) - g\left(E\left(Y^0 \mid X = x\right)\right)$$

Conditional treatment effect at the mean *

$$\text{CTEM} = g\left(E\left(Y^1 \mid X = \bar{X}\right)\right) - g\left(E\left(Y^0 \mid X = \bar{X}\right)\right)$$

*Arguably, not relevant for population-level decision-making

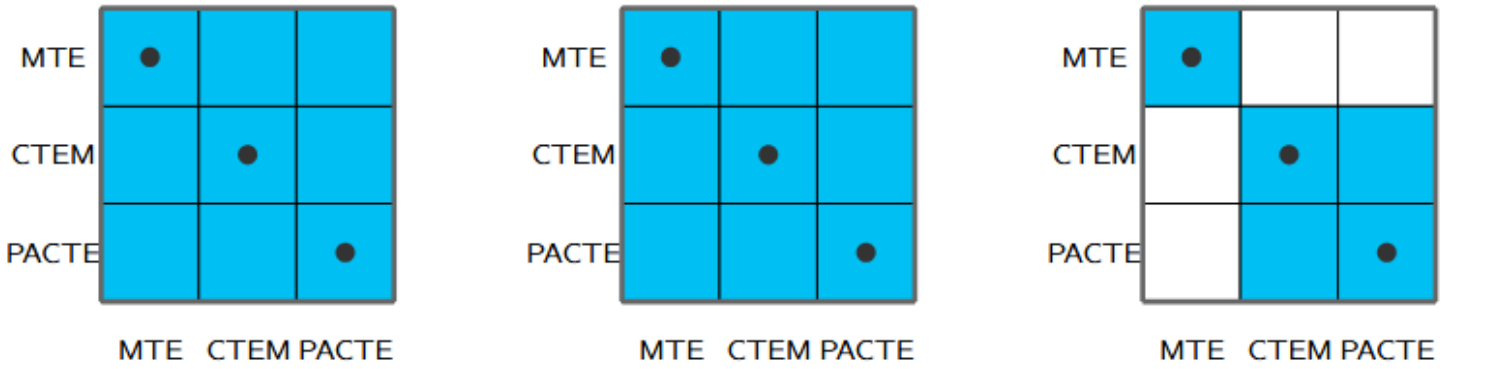
Homogeneous (constant) CATE

Model -based estimands

$$E(Y^t \mid X) = g^{-1}\left(\beta_0 + \beta_X X + \beta_T t\right)$$

Table 2. Model-based marginal estimands for the homogeneous illustrative models. For count outcomes and the log link, person-time is assumed constant, such that the log rate ratio is collapsible and can be interpreted a log risk ratio.

Outcome	Link function	Summary measure	Marginal estimand
Continuous	Identity	Mean difference	Does not depend on the distribution of purely prognostic covariates
Count	Logarithmic	Log risk ratio	Does not depend on the distribution of purely prognostic covariates
Binary	Logit	Log odds ratio	Depends on the full joint distribution of purely prognostic covariates



(a) Identity link, mean difference **(b)** Logarithmic link, log risk ratio **(c)** Logistic link, log odds ratio

Figure 1. Matrices indicating whether different estimands are equivalent for the homogeneous illustrative models. The blue squares denote matching estimand values; the dots denote the diagonal, where estimands are equivalent by definition.

Remiro-Azócar, A., Phillippo, D.M., Welton, N.J., Dias, S., Ades, A.E., Heath, A. and Baio, G., 2025. Marginal and conditional summary measures: transportability and compatibility across studies. arXiv preprint arXiv:2507.21925.

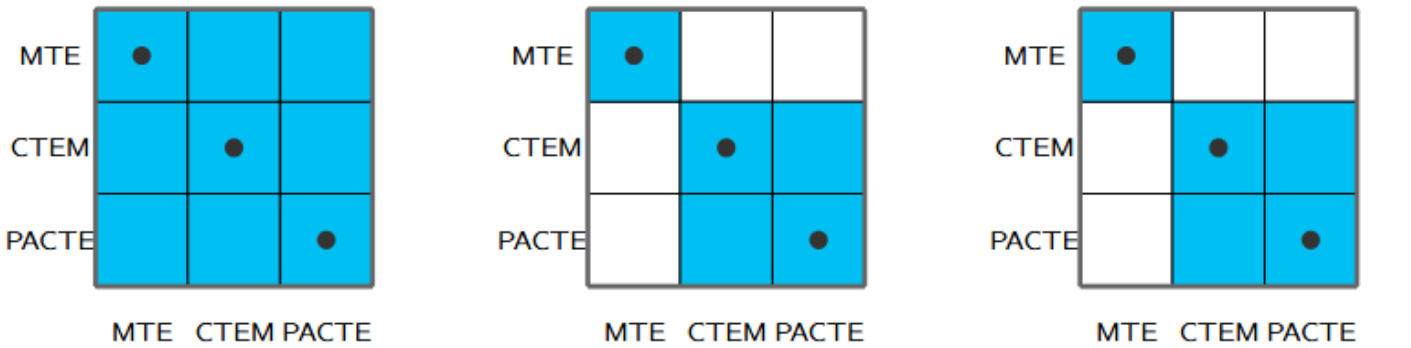
Linear heterogeneous CATE

Model -based estimands

$$E(Y^t \mid X) = g^{-1}\left(\beta_0 + \beta_X X + \beta_T t + \beta_{XT} X t\right)$$

Table 3. Model-based marginal estimands for the heterogeneous illustrative models. For count outcomes and the log link, person-time is assumed constant, such that the log rate ratio is collapsible and can be interpreted as a log risk ratio.

Outcome	Link function	Summary measure	Marginal estimand
Continuous	Identity	Mean difference	Only depends on effect modifier means
Count	Logarithmic	Log risk ratio	Depends on the full joint distribution of effect modifiers and purely prognostic covariates that are associated with the former
Binary	Logit	Log odds ratio	Depends on the full joint distribution of effect modifiers and purely prognostic covariates



(a) Identity link, mean difference **(b)** Logarithmic link, log rate ratio **(c)** Logistic link, log odds ratio

Figure 2. Matrices indicating whether different estimands are equivalent for the heterogeneous illustrative models. The blue squares denote matching estimand values; the dots denote the diagonal, where estimands are equivalent by definition.

Remiro-Azócar, A., Phillippo, D.M., Welton, N.J., Dias, S., Ades, A.E., Heath, A. and Baio, G., 2025. Marginal and conditional summary measures: transportability and compatibility across studies. arXiv preprint arXiv:2507.21925.

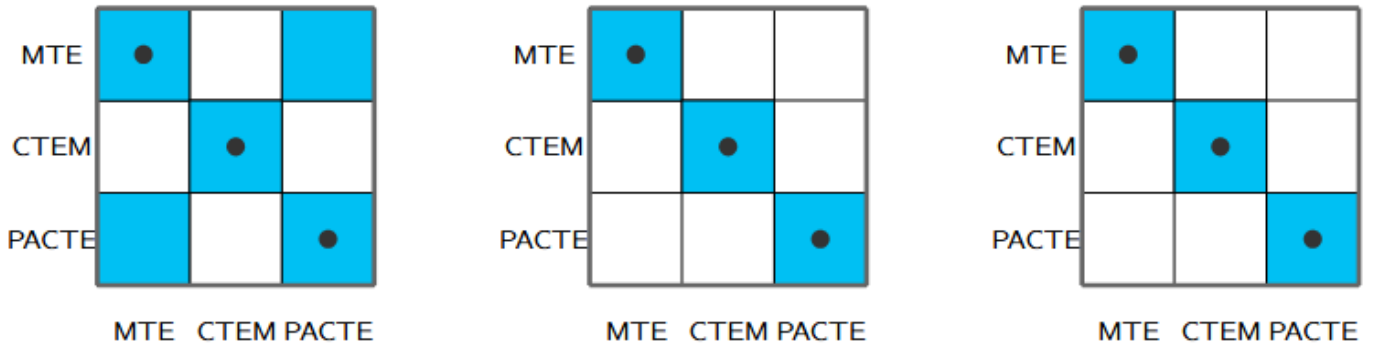
Non-linear (quadratic) heterogeneous CATE

Model -based estimands

$$E(Y^t \mid X) = g^{-1}\left(\beta_0 + \beta_1 X + \beta_2 X^2 + \beta_T t + \beta_{1T} X t + \beta_{2T} X^2 t\right)$$

Table 4. Model-based marginal estimands for the quadratic (heterogeneous) illustrative models. For count outcomes and the log link, person-time is assumed constant, such that the log rate ratio is collapsible and can be interpreted as a log risk ratio.

Outcome	Link function	Summary measure	Marginal estimand
Continuous	Identity	Mean difference	Depends on effect modifier means and variances
Count	Logarithmic	Log risk ratio	Depends on the full joint distribution of effect modifiers and purely prognostic covariates that are associated with the former
Binary	Logit	Log odds ratio	Depends on the full joint distribution of effect modifiers and purely prognostic covariates



(a) Identity link, mean difference **(b)** Logarithmic link, log risk ratio **(c)** Logistic link, log odds ratio

Figure 3. Matrices indicating whether different estimands are equivalent for the quadratic (heterogeneous) illustrative models. The blue squares denote matching estimand values; the dots denote the diagonal, where estimands are equivalent by definition.

Remiro-Azócar, A., Phillippo, D.M., Welton, N.J., Dias, S., Ades, A.E., Heath, A. and Baio, G., 2025. Marginal and conditional summary measures: transportability and compatibility across studies. arXiv preprint arXiv:2507.21925.

What about the hazard ratio?

Even for a constant CATE on the (log) hazard ratio scale…

..the marginal (log) hazard ratio depends on the shape of the hazard function, the distributions of the baseline hazard and observed purely prognostic factors, the length of follow-up and observed censoring pattern!

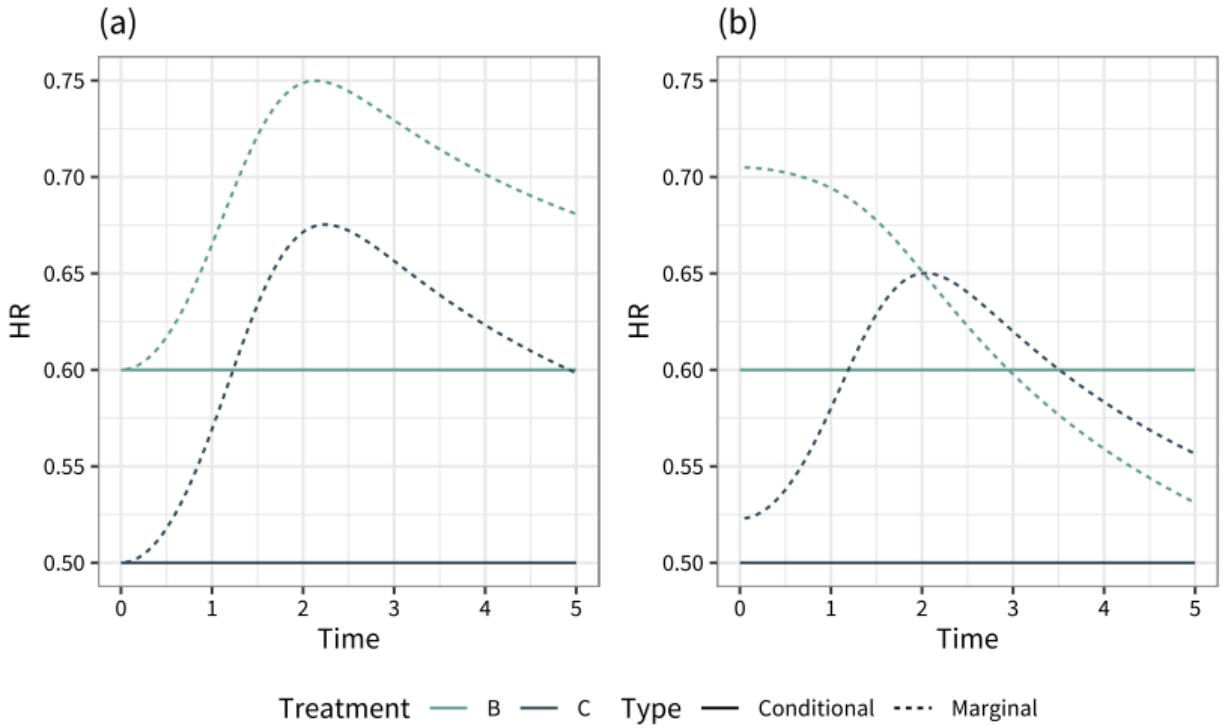


Figure 7. *Population-average conditional and marginal hazard ratios vs. treatment A over time with a single uniformly-distributed covariate that is (a) prognostic only, (b) prognostic and effect modifying.*

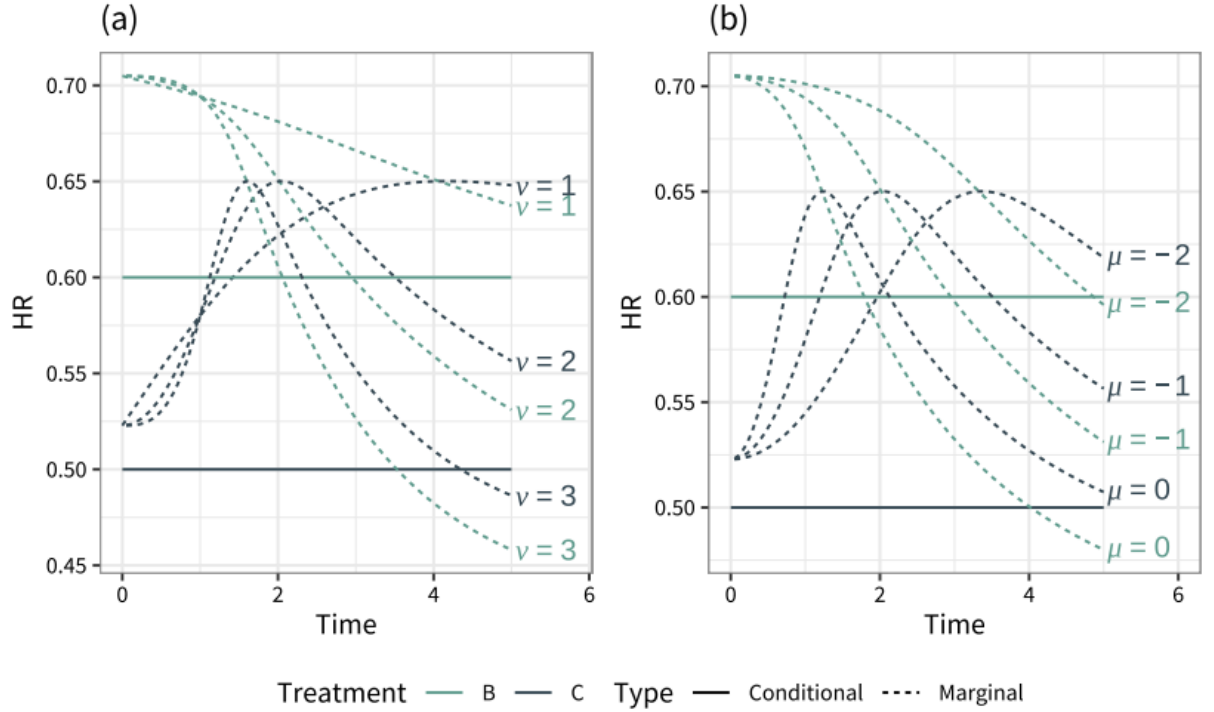


Figure 8. *Population-average conditional and marginal hazard ratios vs. treatment A over time, varying (a) the shape of the baseline hazard function $\nu_{(P)}$, and (b) the distribution of baseline log hazard $\mu_{(P)}$.*

Phillippo, D.M., et al., 2025. Effect modification and non-collapsibility together may lead to conflicting treatment decisions: A review of marginal and conditional estimands and recommendations for decision-making. Research Synthesis Methods, 16(2), pp.323-349.

Marginal versus conditional

Individual trial level

Marginal treatment effect:

- Depends on the distribution of "baseline risk" and observed prognostic factors
- Estimand (summary measure) does not depend on the covariate adjustment set
- Can be identified from RCT data with minimal assumptions, even if adjusting for covariates

Conditional treatment effect:

- Do not depend on the distribution of baseline risk/purely prognostic factors
- Estimand (summary measure) changes with the covariate adjustment set
- Identification may require statistical assumptions about model validity

While within -trial estimation of the marginal estimand may require weaker statistical assumptions, it may require stronger assumptions for transportability across different populations!

Marginal versus conditional

Meta -analytical implications

Meta-analysis of marginal effects (based on aggregate-level data):

- Not valid if populations are heterogeneous
- Only reasonable if the population in the scope is narrowly defined

Meta-analysis of conditional effects (based on aggregate-level data):

- Not valid if summary measures are incompatible, due to conditioning on different covariate sets
- ..may also break down if populations are heterogeneous in observed or unobserved effect modifiers
- Only reasonable if all conditional measures adjust for the same set of covariates

Full access to IPD provides solutions:

- Pool adjusted marginal estimates that have been transported to the same target population
- Meta-regression (targets conditional effect) followed by standardization (targets marginal effect) – can produce any desired marginal or conditional summary measure in any target population

Apples and oranges

The naïve pooling of marginal and conditional summary measures can produce bias:

- Also, for some collapsible measures in the presence of effect modification!
- Trials may employ different analysis methods and report different summary measures in publications

Table 5. Summary measures targeted by selected estimators within RCTs.

Summary measure	Analytical approaches
Marginal	Crude unadjusted difference, simple regression of outcome on treatment, G-computation, IPTW
Conditional	Direct regression adjustment, contrast of least squares means, Mantel-Haenszel methods

Incompatibility issues also problematic for **cross-trial** covariate adjustment methods!

Table 5. Summary measures estimated by different covariate adjustment approaches in the context of indirect comparisons.

Methodology	Summary measure
Matching-adjusted indirect comparison (MAIC)	MTE
"Plug-in" simulated treatment comparison (STC-P)	CTEM
"G-computation" simulated treatment comparison (STC-G)	MTE
Multilevel network meta-regression (ML-NMR)	PACTE and MTE
Network meta-interpolation (NMI)	CTEM
Cross-network meta-regression (cross-NMR)	CTEM

Remiro-Azócar, A., Phillippo, D.M., Welton, N.J., Dias, S., Ades, A.E., Heath, A. and Baio, G., 2025. Marginal and conditional summary measures: transportability and compatibility across studies. arXiv preprint arXiv:2507.21925.

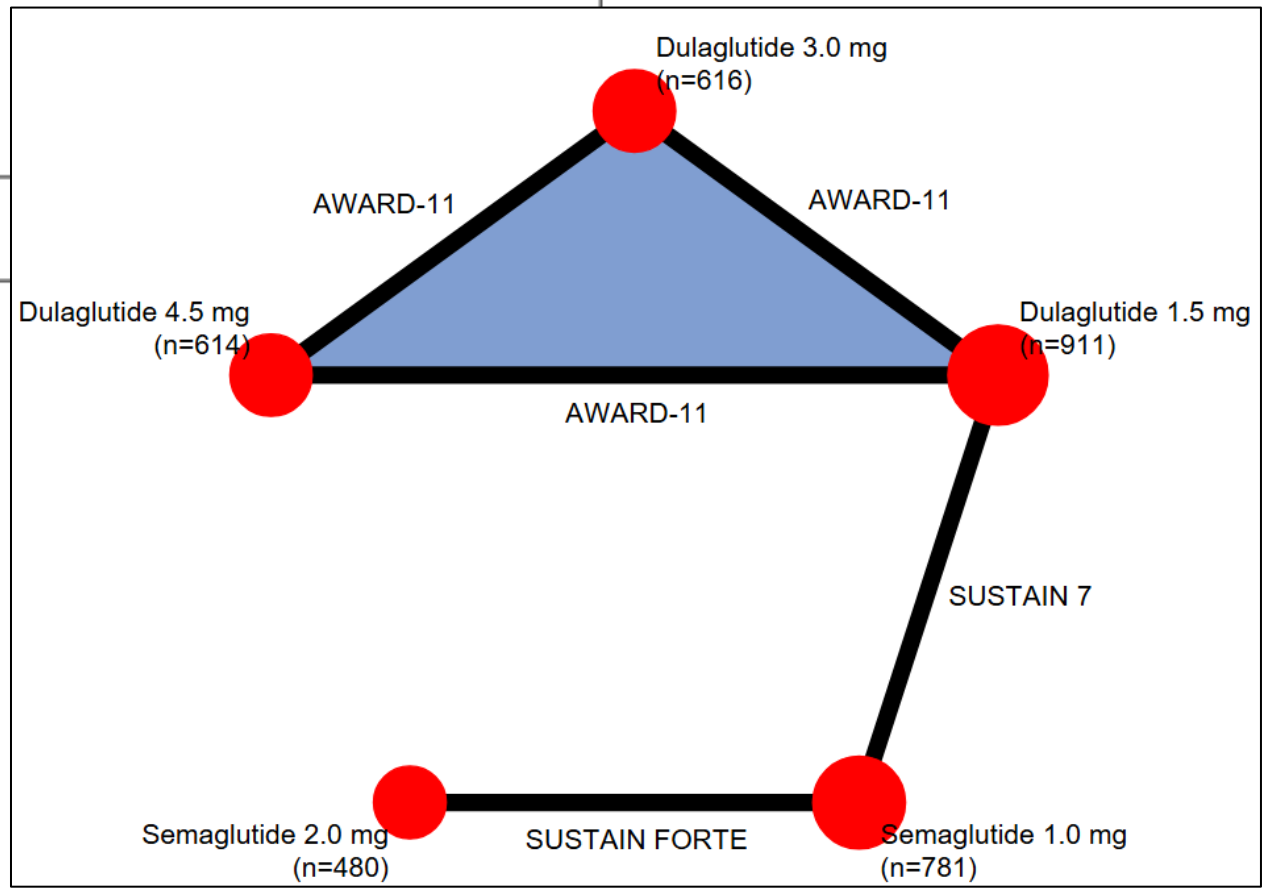
Strategies for intercurrent events

Illustrative example

Population	Subjects with Type 2 diabetes on a background treatment of metformin
Intervention	Semaglutide 2.0 mg QW
Comparators	Dulaglutide 4.5 mg QW Dulaglutide 3.0 mg QW *Dulaglutide 1.5 mg QW* *Semaglutide 1.0 mg QW*
Outcomes	Change from baseline in HbA1c Change from baseline in body weight

Aggregate -level data network meta-analysis comparing semaglutide versus dulaglutide in patients with Type 2 diabetes (T2D)

T2D: therapeutic area motivating the regulatory discussions influencing ICH E9 (R1); RCTs contain information on estimands in trial publications, protocols and SAPs



Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials

Trial-level estimands

According to the full estimand descriptions in the SAPs (wording is somewhat outdated currently):

- All RCTs include a **hypothetical** estimand that characterizes the treatment effect if patients do not discontinue treatment prematurely or initiate anti-diabetic rescue medication

- All RCTs include a **treatment policy** estimand that characterizes the treatment effect regardless of premature treatment discontinuation or initiation of anti-diabetic rescue medication

	Primary estimand	Secondary (FDA-preferred) estimand
AWARD-11	**Efficacy** estimand (de jure effect), which will use the data collected before initiation of any rescue medication or premature treatment discontinuation to demonstrate the effect of treatment and avoid confounding effects of other anti-hyperglycemic agents. The efficacy estimand measures the benefit of treatment when taken as directed.	**Treatment regimen** estimand (de facto effect), which will include data collected after initiation of other anti-hyperglycemic therapy and/or after premature treatment discontinuation. The treatment regimen estimand measures the benefit of treatment as actually taken (that is, irrespective of adherence to investigational product or introduction of other anti-hyperglycemic therapy).
SUSTAIN 7	**De-jure** treatment difference at week 40 for all randomized subjects if all subjects adhered to treatment and did not initiate anti-diabetic rescue medication. This estimand assesses the benefit a future subject is expected to achieve if he/she initiates and continues treatment with subcutaneous semaglutide as compared to dulaglutide.	**De-facto** treatment difference at week 40 for all randomized subjects. This estimand assesses the average effect in a future population that results from treatment with subcutaneous semaglutide plus anti-diabetic rescue medication(s) as compared to treatment with dulaglutide plus anti-diabetic rescue medication(s).
SUSTAIN FORTE	The **hypothetical** estimand, which is the absolute treatment difference in mean change from baseline to week 40 of semaglutide 2.0 mg versus semaglutide 1.0 mg, both as an add-on to metformin with or without sulphonylurea, in all randomized subjects with T2D, regardless of change in treatment dose and had they not discontinued treatment due to adverse events or initiated any rescue medication (anti-diabetic medications).	The **treatment policy** estimand for the primary objective will be estimated as the absolute treatment difference in mean change from baseline to week 40 of semaglutide 2.0 mg versus semaglutide 1.0 mg, both as an add-on to metformin with or without sulphonylurea, in all randomized subjects with T2D, regardless of change in treatment dose, discontinuation of treatment due to adverse events and initiation of rescue medication (anti-diabetic medications).

Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials

Trial-level estimands

Trial	AWARD-11[60,66,67]		SUSTAIN 7[57,62,63]		SUSTAIN FORTE[59,64,65]	
Estimand name	Efficacy	Treatment regimen	De-jure	De-facto	Hypothetical (trial product)	Treatment policy
Population	Subjects with Type 2 diabetes inadequately controlled with metformin		Subjects with Type 2 diabetes on a background treatment with metformin		Subjects with Type 2 diabetes on a background treatment of metformin with or without sulphonylurea treatment	
Treatments	Subcutaneous dulaglutide 4.5 mg QW Subcutaneous dulaglutide 3.0 mg QW Subcutaneous dulaglutide 1.5 mg QW		Subcutaneous semaglutide 1.0 mg QW Subcutaneous semaglutide 0.5 mg QW Subcutaneous dulaglutide 1.5 mg QW Subcutaneous dulaglutide 0.75 mg QW		Subcutaneous semaglutide 2.0 mg QW Subcutaneous semaglutide 1.0 mg QW	
Variables (endpoints)	Primary: change from baseline to week 36 in HbA1c (%-points) Key secondary: change from baseline to week 36 in body weight (kg)		Primary: change from baseline to week 40 in HbA1c (%-points) Confirmatory secondary: change from baseline to week 40 in body weight (kg)		Primary: change from baseline to week 40 in HbA1c (%-points) Confirmatory secondary: change from baseline to week 40 in body weight (kg)	
Population-level summary measure	Mean difference in change from baseline		Mean difference in change from baseline		Mean difference in change from baseline	
Intercurrent event strategies	**Hypothetical** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation	**Treatment policy** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation	**Hypothetical** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation	**Treatment policy** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation	**Hypothetical** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation **Treatment policy** strategy for change in treatment dose	**Treatment policy** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation **Treatment policy** strategy for change in treatment dose

Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials

Can we define an estimand at the ”meta” level?

We can construct two target meta-analytical estimands: one which adopts a hypothetical strategy for premature treatment discontinuation or initiation of anti-diabetic rescue medication and another which adopts a treatment policy strategy for the corresponding intercurrent events

Estimand name	Hypothetical	Treatment policy
Population	Subjects with Type 2 diabetes on a background treatment of (i.e., inadequately controlled with) metformin	
Treatments	Semaglutide 2.0 mg QW Dulaglutide 4.5 mg QW Dulaglutide 3.0 mg QW *Dulaglutide 1.5 mg QW* *Semaglutide 1.0 mg QW*	
Variables (endpoints)	Change from baseline to week 36 or week 40 in HbA1c (%-points) Change from baseline to week 36 or week 40 in body weight (kg)	
Population-level summary measure	Mean difference in change from baseline	
Intercurrent event strategies	**Hypothetical** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation	**Treatment policy** strategy for initiation of anti-diabetic rescue medication or premature treatment discontinuation

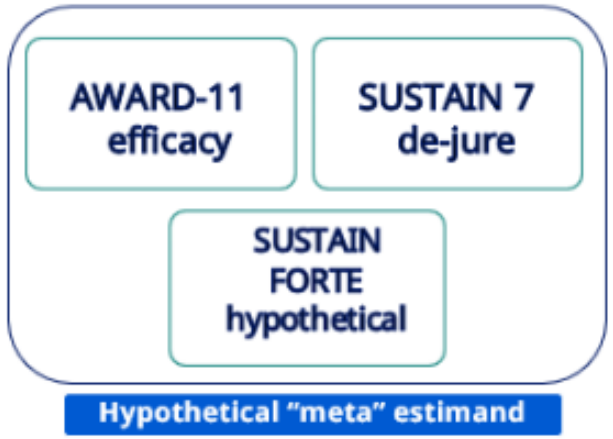
Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials

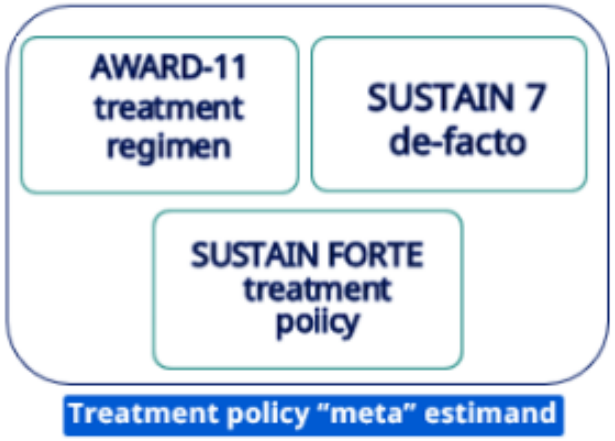
Can we define an estimand at the ”meta” level?

The original PICO scope can be restricted using each meta-analytical estimand to produce more targeted meta-analyses that maximize relevance relative to the healthcare decision-making context

PICO scope

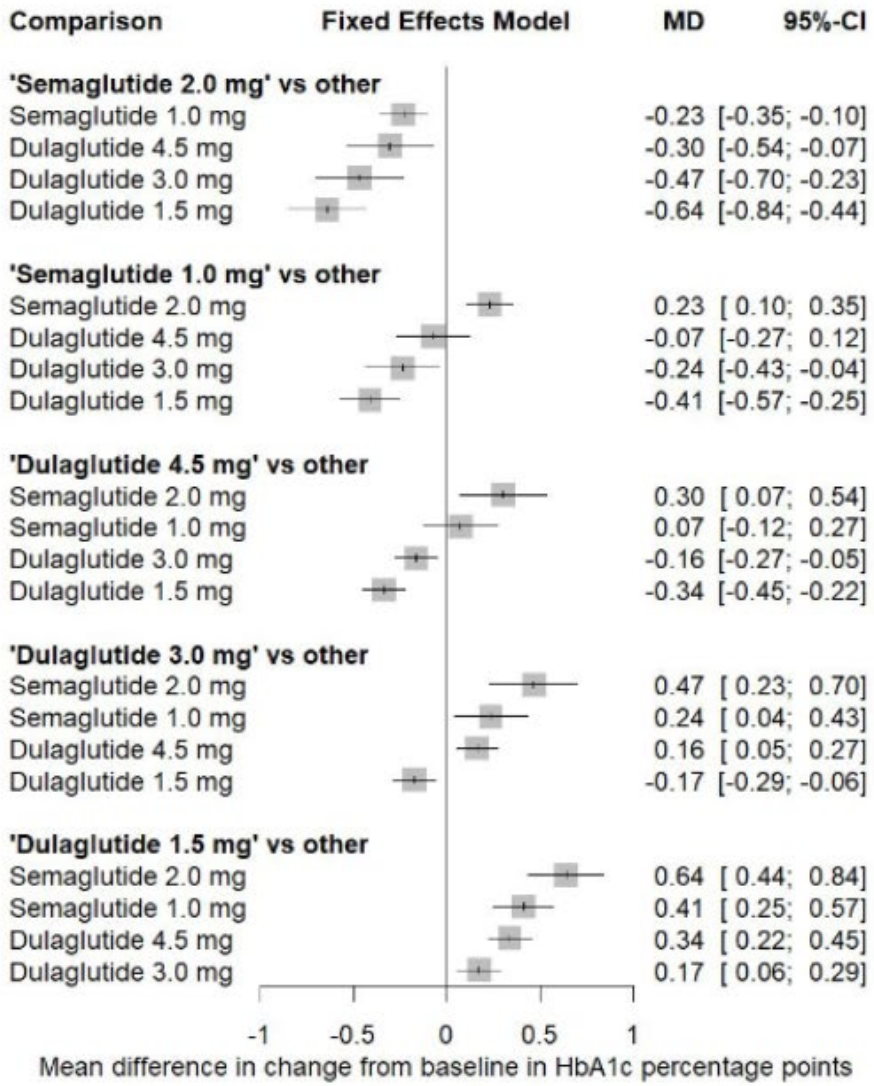
- **P**opulation: subjects with Type 2 diabetes on a background treatment of metformin
- **I**ntervention: semaglutide 2.0 mg
- **C**omparators: dulaglutide 4.5 mg, dulaglutide 3.0 mg, *dulaglutide 1.5 mg*, *semaglutide 1.0 mg*
- **O**utcomes: change from baseline in HbA1c and body weight



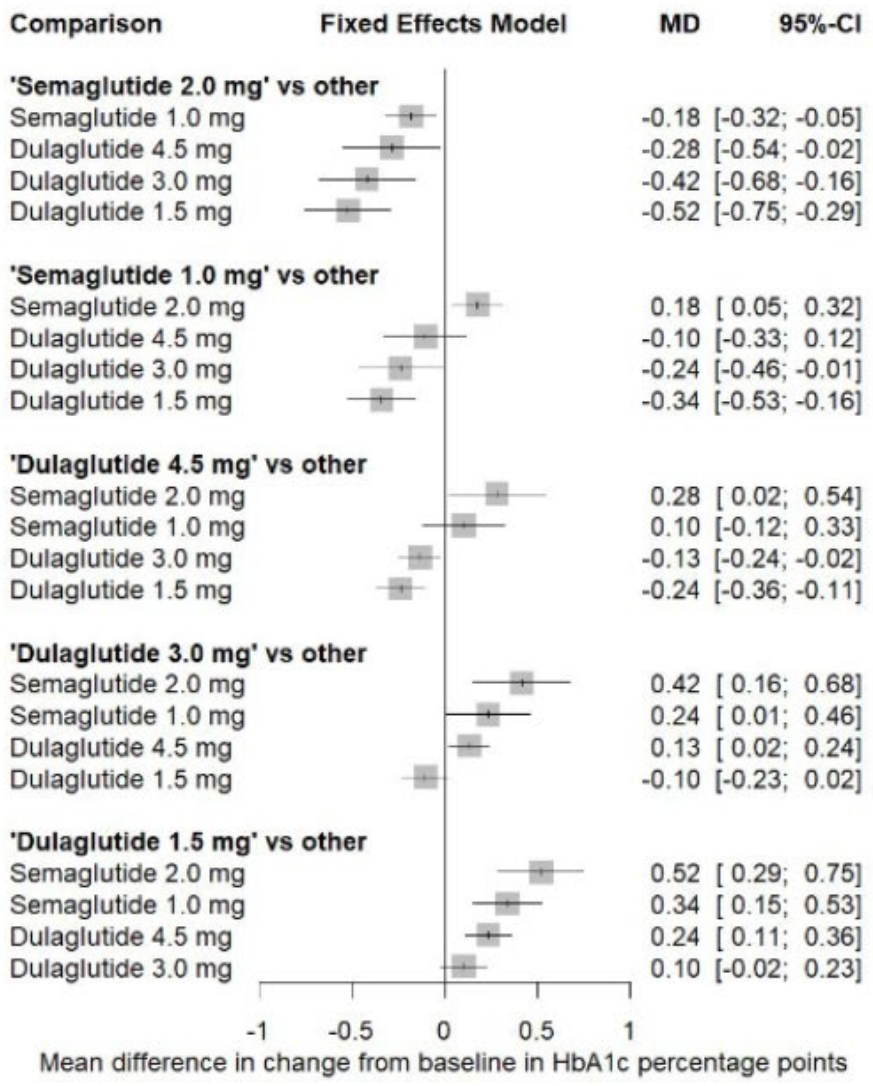


Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials

Results – change from baseline in HbA1c



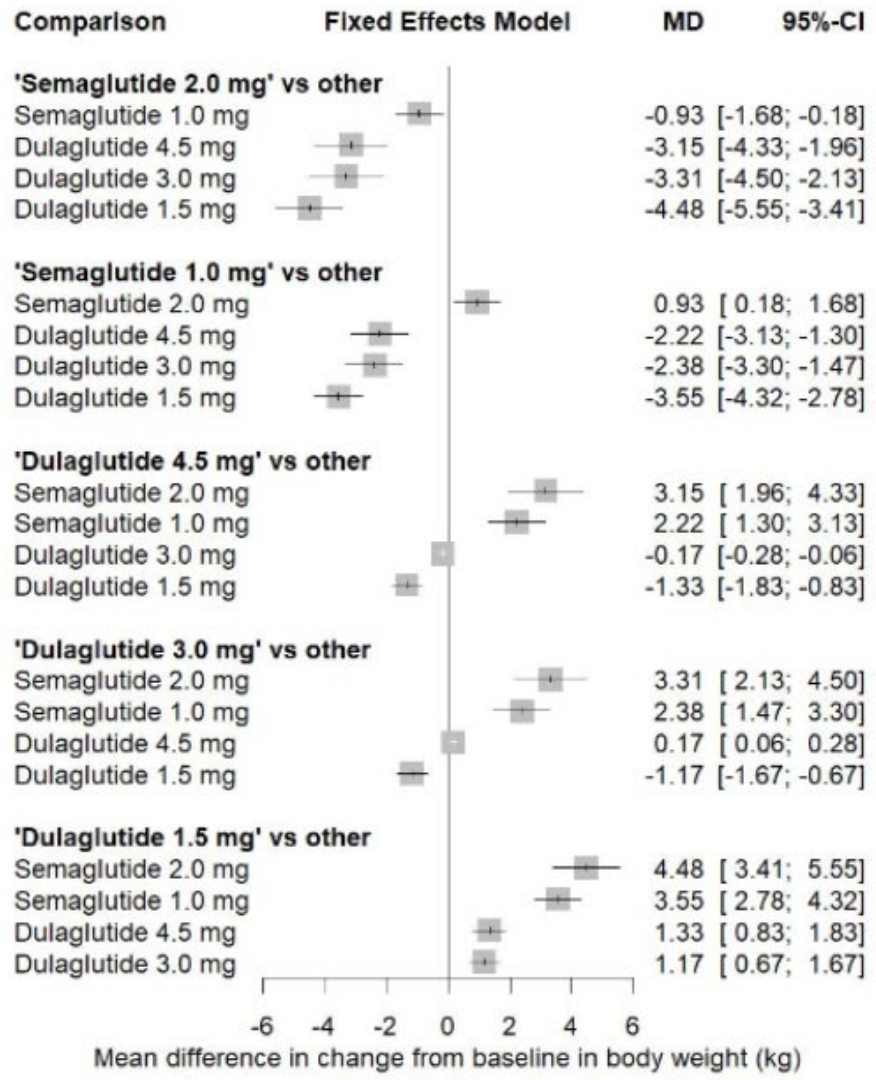
(a) Hypothetical meta-analytical estimand



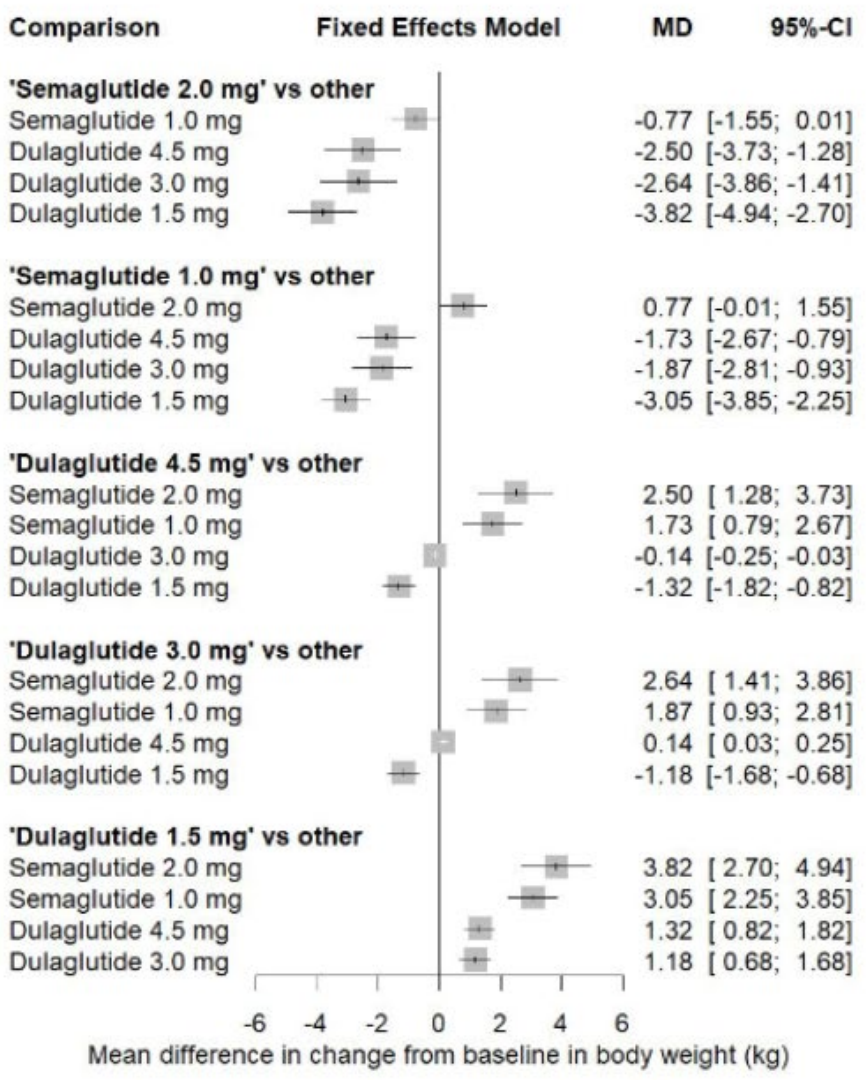
(b) Treatment policy meta-analytical estimand

Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials

Results – change from baseline in body weight



(a) Hypothetical meta-analytical estimand



(b) Treatment policy meta-analytical estimand

Remiro-Azócar, A. et al, 2025. Incorporating estimands into meta-analyses of clinical trials

Discussion

External validity is key for HTA: what estimand reflects treatment pathways in routine clinical practice?

- Treatment policy for treatment discontinuation: "real-world" patients do discontinue treatment
- Hypothetical for initiation of rescue medication: "real-world" patients typically switch within the same drug class, but this is not permitted in comparative GLP-1 RA trials
- A "hybrid" estimand is recommended by EMA

Consistent estimands/intercurrent event strategies are less likely in less "mature" therapeutic areas

Full access to individual patient data would allow to:

- Target the "hybrid" estimand recommended by EMA – at the trial and meta-analytical level
- Pursue alignment between the timing of outcome assessments – note visits did not coincide here
- Perform meta-regression, which can resolve heterogeneity due to population differences

Concluding remarks

Summary measure and intercurrent event strategies:

- At the heart of ICH E9 (R1) but a secondary consideration in evidence synthesis..not components of PICO!
- Overlooked source of heterogeneity in meta-analyses
- Fundamental components of any well-defined clinical research question

Estimands:

- Align with a "narrow" perspective of research questions, required for healthcare decision-making
- A language used in trial design and reporting, which will become commonplace in publications

The uptake of estimands is as important for the synthesis of multiple RCTs as it is for single trials!