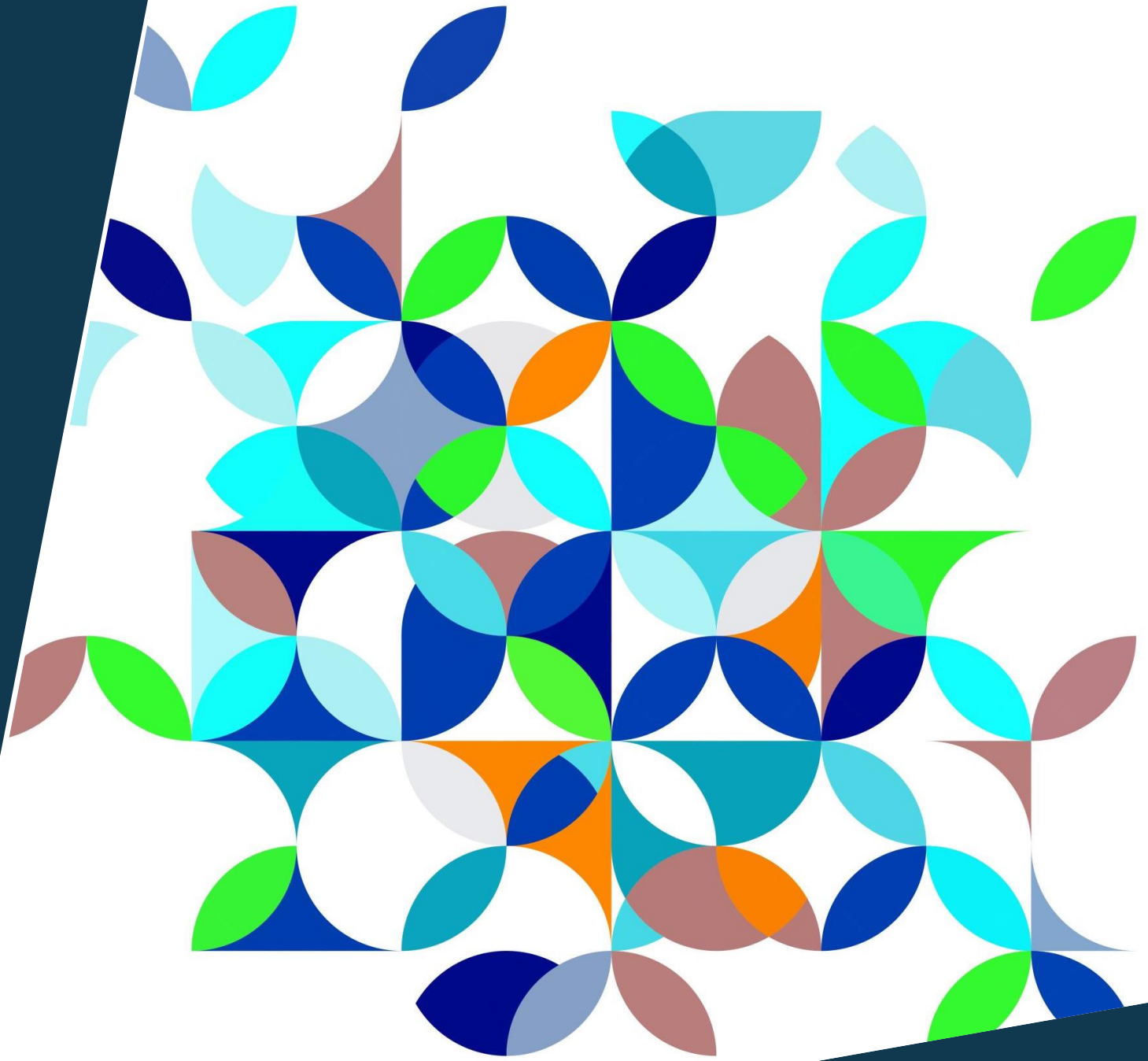
PSI HTA SIG WEBINAR

# *Estimands, PICOs and Co. - Are we losing or gaining in translation?*

**Antonio Remiro-Azócar, PhD**
**Statistics and Data Insights, Bayer**

December 7, 2023



# Disclaimers

- The findings and conclusions in this presentation are those of the presenter, who is responsible for its contents. The findings and conclusions do not necessarily represent the views of Bayer. No statement in this presentation should be construed as an official position of Bayer.

- This presentation has very little statistics

# Structure

1. HTA and the PICO framework
2. PICO in the context of the EU HTA Regulation
3. The inherent tension between broad and narrow research questions
4. Estimands in HTA: opportunities and challenges

There will be a specific focus on evidence synthesis

# Health technology assessment (HTA)

What is **health technology assessment**?

- A **health technology** is an intervention used to promote health, e.g., a pharmaceutical product or a medical device
- An **assessment** is required to inform policy decision-making
- The assessment is **multidisciplinary**, involving clinical, social, economic, organizational and ethical aspects

# Regulatory versus HTA statistics in drug development

## REGULATORY

- Focus on safety, clinical efficacy, quality, benefit-risk
- Focus on hypothesis testing
- **Estimands**
- Time-horizon is the trial follow-up; does not typically require extrapolation
- Relies mostly on data of a “pivotal” Phase III clinical trial as the primary source of evidence
- Direct comparator is usually placebo or standard of care in a head-to-head study

## HTA

- Focus on the “fourth hurdle”: clinical effectiveness (value) and cost-effectiveness (value-for-money)
- Focus on estimation
- **PICOs**
- Long-term or “lifetime” horizon; may require extrapolation beyond the trial follow-up
- Typically requires secondary data sources beyond the “pivotal” clinical trial
- Comparators are all competing treatment options; direct comparisons may be unavailable

# The PICO framework

In HTA, the **PICO** framework is used to translate national policy questions into research questions. Relevant PICO question(s) are specified in the HTA scoping process. Manufacturers submit an evidence dossier to HTA agencies addressing the PICO question(s) in the scope.

PICO questions are sometimes supplemented by additional criteria such as study type (PICOS) or length of follow-up/duration of study (PICOT)

PICO Concept Paper (EUnetHTA Joint Action 3):

***3.2 The PICO framework***

The PICO framework provides a standard format for the definition of a research question, e.g. for a comparative assessment of the effectiveness and safety of various treatment options.

Within the PICO framework research questions are defined using (at minimum) the following components:

| | |
|---|---|
| P (population) | the patients or population(s) in which the intervention under assessment should be used |
| I (intervention) | the therapeutic, diagnostic or preventive intervention under assessment (incl setting) |
| C (comparator) | the alternative intervention(s) against which the intervention under assessment should be compared |
| O (outcomes) | the outcomes of interest (if relevant incl. minimum follow-up time) |

# PICO example

NICE final scope to appraise the effectiveness of cabozantinib within its marketing authorization for previously treated locally advanced or metastatic differentiated thyroid cancer (TA928)

| | |
|---|---|
| **Intervention** | 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. |
| **Comparator** | Best supportive care |
| **Outcomes** | The outcome measures to be considered include:<br>• overall survival<br>• progression-free survival<br>• response rate<br>• adverse effects of treatment<br>• health-related quality of life. |

# PICOs in EUnetHTA

PICO Concept Paper (EUnetHTA Joint Action 3):

- “The starting point for an assessment of a medical intervention is the formulation of a defined research question to be investigated, based on a policy question raised by healthcare decision makers”
- “The PICO framework provides a standard format for the definition of a research question and helps to specify the data requirements for the assessment. The PICO is specified prior to initiation of any assessment.”
- “A clear definition of the PICO question is required to guide the development of a joint assessment”
- “Because of different policy questions from different partners or because of different research questions within the complete approved indication of a specific treatment, it is possible that more than one PICO is required to define the research questions to be answered in a given assessment.”

# The EU HTA Regulation (EUnetHTA21)

The assessment scope (PICO) is determined by the EU member states, before being consolidated by the “Coordination Group”

The scoping process can lead to the inclusion of multiple PICOs in trying to meet the needs of all states:

- Local/historical variations in routine clinical practice across member states
- Differences in relevant comparators across countries, e.g., standard of care at time of launch (potentially including non-EMA authorized treatments)
- Differences in approved indications, targeted populations, e.g., different lines of therapy
- Particularly in rapidly evolving therapeutic areas such as oncology

EUnetHTA 21 Practical Guideline D4.2 Scoping Process

**Table 3-8: Consolidated PICOs based on Member State requests**

| | PICO 1 | PICO 2 | PICO 3 | PICO 4 | PICO 5 |
|---|---|---|---|---|---|
| P | Full licensed indication | Full licensed indication | Full licensed indication | Subpopulation A | Subpopulation B |
| C | Comparator 1 OR Comparator 2[14] | Comparator 3 | Comparator 4 | Comparator 1 | Comparator 3 |
| O | All outcomes | All outcomes | All outcomes | All outcomes | All outcomes |

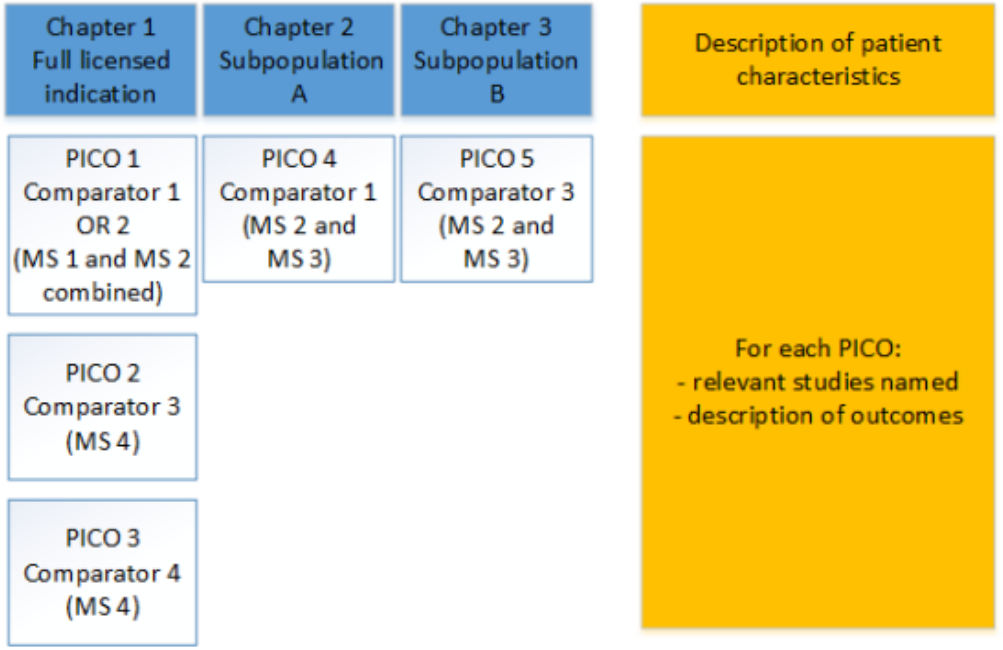



**Figure 5-1: Data presentation according to PICO(s).**

MS, Member State; PICO=Population, Intervention, Comparators, Outcomes.

# A multitude of PICOs

## Consolidated PICO exercise for Pombiliti (EUnetHTA 21): 10 EU member states participated

*Table 2 - Parameters for the assessment scope of Pombiliti (1 PICO per comparator)*

| Description of PICO elements | P | I | C | O |
|---|---|---|---|---|
| PICO 1 | Full Population: Adult patients with late onset Pompe disease | Cipaglucosidase alfa in combination with miglustat | Alglucosidase alfa | See outcomes table |
| PICO 2 | as for PICO 1 | as for PICO 1 | Avalglucosidase alfa | |
| PICO 3 | as for PICO 1 | as for PICO 1 | BSC | |
| PICO 4 | as for PICO 1 | as for PICO 1 | Physicians choice for control arm, with at least:<br>- Alglucosidase alfa<br>- Avalglucosidase alfa | |
| PICO 5 | Subpopulation: Adult patients with late onset Pompe disease, who are ERT-naive | as for PICO 1 | Alglucosidase alfa | |
| PICO 6 | as for PICO 5 | as for PICO 1 | Avalglucosidase alfa | |
| PICO 7 | as for PICO 5 | as for PICO 1 | BSC | |
| PICO 8 | Subpopulation: Adult patients with late onset Pompe disease, who are ERT-experienced | as for PICO 1 | Alglucosidase alfa | |
| PICO 9 | as for PICO 8 | as for PICO 1 | Avalglucosidase alfa | |

*Abbreviations: BSC: Best supportive care; ERT: Enzymatic replacement therapy.*

**Outcomes table**

| |
|---|
| Overall survival |
| Ventilator-free survival |
| Changes in mobility (incl. measurement by 6MWT and documented use of wheelchair) |
| Changes in respiratory function (incl. measurement by FVC in sitting and upright positions) |
| Changes in muscle strength (by validated scales) |
| Changes in motor function (by validated scales, e.g. quick motor function test) |
| Respiratory symptomatology associated with Pompe disease |
| Gastrointestinal symptomatology associated with Pompe disease |
| Quality of life (as assessed using disease-specific (preferably) and/or generic questionnaires) |
| Health status (measured preferably by the EQ-5D) |
| Patient-reported outcomes to include R-PAct scale, and any other patient-centered outcome assessed by means of a patient-reported outcome measure |
| Adverse events (AEs) (incl. hypersensitivity, infusion reactions, immunogenicity) |
| Serious AEs (SAEs) |
| Severe AEs |
| Discontinuation and interruption of treatment due to AEs |
| Mortality due to AEs |

*Abbreviations: 6MWT: 6-minute walking test; AE: Adverse event; EQ-5D: EuroQoL five-dimension scale questionnaire; FVC: Forced vital capacity; R-PAct: Rasch-built Pompe-specific Activity.*

- Some misalignment in PICO elements between states
- The full population is split into narrower subpopulations
- Comparators are broad or more defined
- The pivotal RCT (Pombiliti vs. alglucosidase alfa vs. placebo) does not directly compare the intervention with all relevant comparators
- Requests for analyses that have not been pre-specified in the pivotal RCT
- The endpoints in the outcomes table are relatively broad
- With 1 PICO per outcome, PICOs would grow considerably

# Three levels of PICO

PICOs originate in systematic literature reviews, evidence-based medicine and evidence synthesis

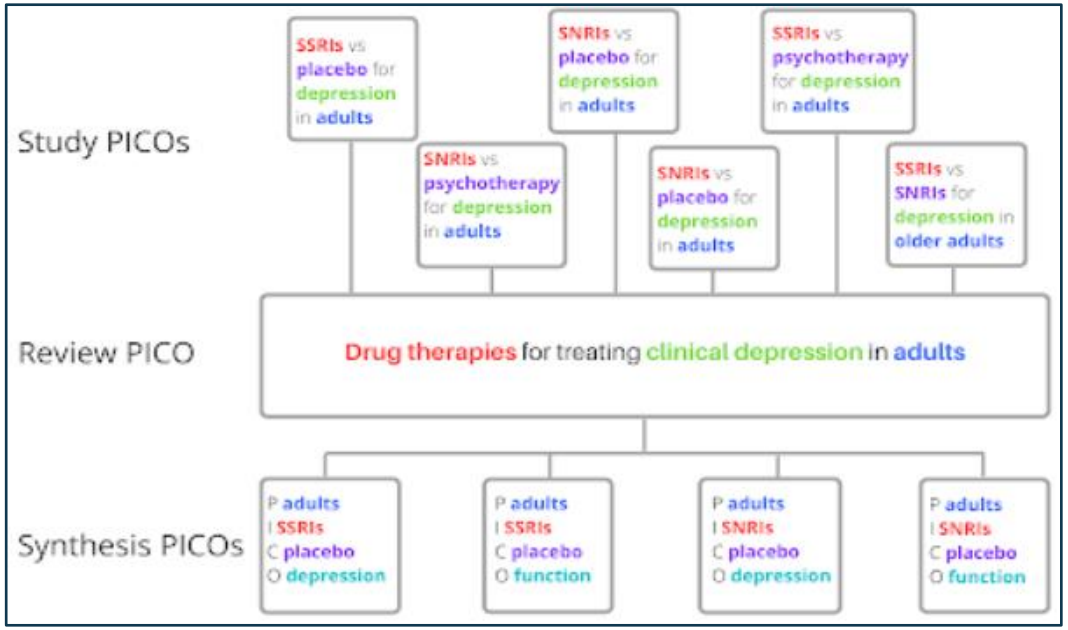
PICOs have traditionally formulated **broad** research questions:

- To guide the data extraction process required for a review
- To keep the evidence base reasonably large and enable the synthesis of many studies

Nevertheless, PICOs can be refined to address more specific research questions

**Three levels of PICO** (Cochrane Handbook for Systematic Reviews of Interventions):

1. **Review PICO**: determines the boundary for study eligibility in a systematic literature review
2. **Synthesis PICOs:** research questions for specific (direct or indirect) treatment comparisons
3. **Study PICOs**: trial-level questions investigated in individual studies (estimands?)




https://www.covidence.org/blog/pico-all-you-need-to-know/

# Broad versus narrow synthesis PICOs

Advantages of broad PICOs for treatment comparisons:

- Less fragmented evidence base; more data, lower risk of too few studies
- Greater likelihood of complete evidence networks
- (Potentially) greater alignment in PICO elements across states
- Lower burden of analytical complexity, e.g., in indirect treatment comparisons
- Sparse evidence base more likely to reflect proprietary or individual interests (?)

Disadvantages:

- Inconsistent set of populations, comparators, outcomes
- “Apples and oranges”: inappropriate mixing of studies within treatment comparisons
- Questionable scientific legitimacy, particularly where the extent of effect modification and heterogeneity is large
- A PICO that is too imprecise leads to GIGO

# Broad versus narrow synthesis PICOs

While some level of aggregation might be necessary for an evidence synthesis to be conducted, the uncritical aggregation of PICO components may threaten the validity of the analysis

**Population**

- Studies may have different target populations due to different selection criteria
- Differences between samples/analysis sets also arise due to sampling variability, non-random sampling
- “Random effects” analyses may be used to account for heterogeneity but do not produce estimates in a specific sample/population

**Intervention/Comparator(s)**

- Potential dissimilarities between the treatments and versions of treatment(s) being administered in different studies, e.g., dosing formulation, delivery mechanism, co-treatment regimens, concomitant background medications

**Outcome(s)**

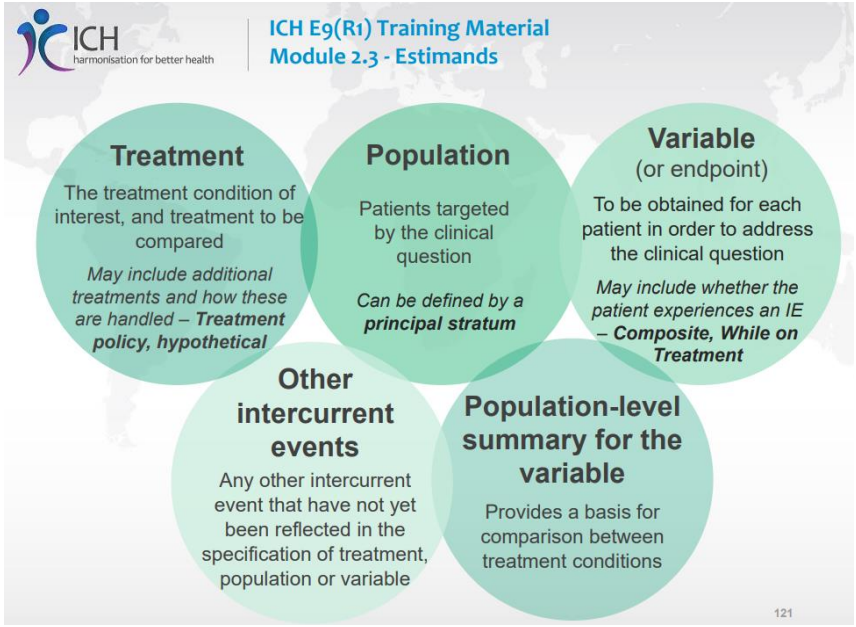
- Differences in endpoints measured using different methods, at different times, or in studies with different durations/follow-up times
- Standardization across studies limited by unavailable subject-level data for competitor studies

# Estimands

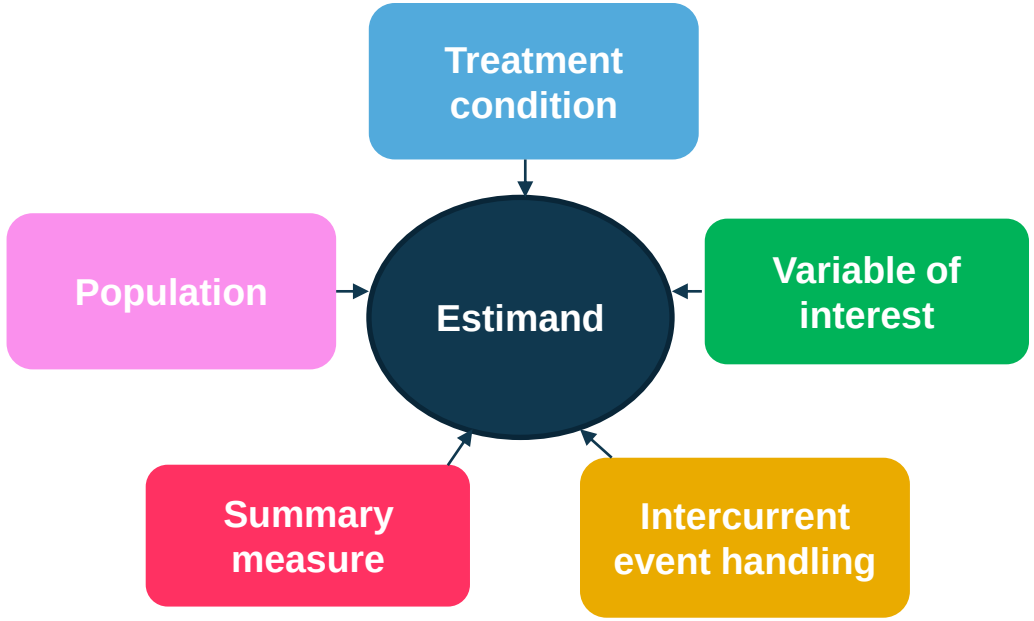
For RCTs in the regulatory setting, the use of **estimands** has been stimulated by the publication of the ICH E9 (R1) Addendum, recognized by the FDA and the EMA

Estimands address very precise research questions (arguably more precise than the most precise of PICOs), defined by the study team at the study level

According to the ICH guidance, estimands should be defined based on the following attributes:




https://database.ich.org/sites/default/files/E9%28R1%29%20Training%20Material%20-%20PDF_0.pdf



# Estimand example

Primary estimands for COSMIC-311, the “pivotal” RCT used to obtain marketing authorization for cabozantinib for previously treated locally advanced or metastatic differentiated thyroid cancer

**Table 7. Primary Estimand Attribute for ORR**

| Estimand attribute [1] | Primary definition for study | |
|---|---|---|
| Population | Subjects randomized into the study intended to include patients with radioiodine-refractory differentiated thyroid cancer who have progressed after prior VEGFR-targeted therapy. | |
| Endpoint | Radiographic response per RECIST 1.1 | |
| Intercurrent events | **Event** | **Strategy** |
| | Receipt of assigned study treatment | Treatment policy |
| | Receipt of local radiation to bone | Treatment policy |
| | Surgical resection of non-target tumor lesions | Treatment policy |
| | Death | Treatment policy |
| | Loss to radiographic follow up | Treatment policy |
| | Receipt of local non-protocol anti-cancer medications other than for disease under study | Treatment policy |
| | Surgical resection of target tumor lesions | While on treatment* |
| | Receipt of systemic non-protocol anti-cancer medications | While on treatment* |
| | Receipt of local non-protocol anti-cancer medications for disease under study | While on treatment* |
| | Receipt of local radiation to soft tissue for disease under study | While on treatment* |
| Population summary | 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. | |
| Estimator | Fisher’s exact test | |

* A modified version of the “while on treatment” strategy is employed for these intercurrent events. Only data prior to the occurrence of these intercurrent events is of interest, but under the ITT principle, receipt of study treatment itself is not considered.

**Table 8. Primary Estimand Attribute for PFS**

| Estimand attribute [1] | Primary definition | |
|---|---|---|
| Population | Subjects randomized into the study intended to include patients with radioiodine-refractory DTC who have progressed after prior VEGFR-targeted therapy. | |
| Endpoint | Duration of radiographic progression-free survival | |
| Intercurrent events | **Event** | **Strategy** |
| | Receipt of assigned study treatment | Treatment policy |
| | Clinical deterioration | Treatment policy |
| | Receipt of local radiation to bone | Treatment policy |
| | Surgical resection of non-target tumor lesions | Treatment policy |
| | Receipt of local non-protocol anti-cancer medications other than for disease under study | Treatment policy |
| | Surgical resection of target tumor lesions | Hypothetical |
| | Receipt of systemic non-protocol anti-cancer medications | Hypothetical |
| | Receipt of local non-protocol anti-cancer medications for disease under study | Hypothetical |
| | Receipt of local radiation to soft tissue for disease under study | Hypothetical |
| Population summary | Difference in survival functions between treatment conditions. | |

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

# Estimands in HTA

The PICO framework does not:

- Discuss the population-level measure used to summarize treatment effects
- Include a strategy for intercurrent events, at the heart of the ICH E9 (R1) Addendum

In HTA, the term “estimand” is rarely used:

- Do PICOs require refinement to more precisely characterize the inferential target in HTA?
- Should HTAs research questions be mapped into a formal estimand?
- Should HTA stakeholders adopt regulatory language instead of traditional HTA terminology?

**Estimands in the context of the EUnetHTA 21 methodological guidelines**:

- Estimands are mentioned in D4.3.1 Direct and Indirect Comparisons
- D4.5 Applicability of Evidence: “different HTAs could also prefer different estimands. Such differences in preference will be assessed during the scoping process.”
- D4.6 Validity of Clinical Studies: “assess evidence with an analysis strategy that best corresponds to given PICO question (…) as defined according to the principles of the estimand framework”

# The summary effect measure in HTA

The **population-level summary measure** is often a secondary consideration when postulating HTA research questions, when compared to patient-level outcomes or endpoints (EUnetHTA 21 D4.4 Outcomes Endpoints)

However, it is important for aspects that are central to HTA:

**Marginal versus conditional effect measures**

- Marginal estimands are more relevant for population-level decision making
- Combining marginal and conditional estimates in evidence synthesis produces bias for non-collapsible measures; non-collapsible measures are widespread in HTA and often pooled in evidence syntheses

**Relative versus absolute estimands**

- Relative estimands (treatment effects) are the logical choice for comparative effectiveness
- Mean absolute outcomes are often input to health economic decision models

**Evaluation of effect modification**

- Necessary to assess external validity, generalizability, transportability
- Effect modifier status is directly tied to the scale used to measure the summary effect, whether this is marginal or conditional, collapsible or non-collapsible…

# Current uptake of estimands in HTA

## HTA Agencies' awareness of the ICE E9(R1) Addendum

| National Institute for Health and Care Excellence | Institute for Quality and Efficiency in Health Care |
|---|---|
| NICE | IQWiG |
| ☐ Departure from ITT principle in the NICE Health Technology Evaluation manual | ☐ Comments during consultation phase on the ICH E9 (R1) Addendum |
| ☐ No explicit mention to the Addendum in the manual | ☐ Only two strategies (treatment policy, composite) should be used as the main analysis |
| ☐ Adjustment for treatment swithching possible, correcting the confounding effects of the ITT principle | ☐ The other strategies are useful only as a possible supplementary or sensitivity analysis |
| ☐ Technical paper elaborates on bias of possible estimation methods, not on appropriateness of clinical question | ☐ Validity and utility of the hypothetical strategy questioned |

Morga, Gorst-Rasmussen, Polavieja et al. (2023) "Estimands in Health Technology Assessments: methodological considerations and recommendations"

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

<19.02.2018>

Submission of comments on 'ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials' (EMA/CHMP/ICH/436221/2017)

**Comments from:**

| Name of organisation or individual |
|---|
| Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany |

- To date, the ICH E9 (R1) Addendum is mostly not acknowledged by HTA agencies
- There are divergent positions on the relevance of certain intercurrent event strategies for HTA objectives
- Acceptance of hypothetical strategy depends on tolerance for extrapolation beyond the trial follow-up (reliance on mean survival estimates and cost-effectiveness analyses)

# Estimands: implementation challenges in HTA

Regulatory decisions are generally made prior to HTA decisions:

- Pivotal RCTs are typically designed in the premarketing authorization setting
- HTA objectives have traditionally played a limited role in the design of such RCTs

Regulatory and HTA decision-makers have different perspectives:

- Regulatory estimands may not align with the estimands required for HTA
- Internal versus external validity; “efficacy-effectiveness” gap of estimand choices (efficacy in ideal and controlled circumstances versus effectiveness in pragmatic “real-world” conditions)

Multiplicity of stakeholders:

- Different HTA agencies may target different questions, decision problems, estimands
- External validity depends on the question targeted by the payer; applicability will differ between jurisdictions

# Estimands: implementation challenges in HTA

Many different HTA agencies with different requirements…“we should avoid focusing exclusively on the viewpoint of a single stakeholder when designing studies” (Schiel 2022, DOI: 10.1002/sim.9517)

It is difficult to design a trial that meets all stakeholders’ research questions of interest

# Estimands: challenges in evidence synthesis

Common scenario in HTA:

- An evidence synthesis (indirect treatment comparison, meta-analysis) is required for HTA
- This combines the results of multiple RCTs, completed at the time of the evidence synthesis
- Each RCT has been designed for regulatory approval and has target estimands of its own

**“Individual studies were not planned with similar estimands nor were they necessarily planned in anticipation of a meta-analysis”** (Russek-Cohen 2022, DOI: 10.1002/sim.9533)

Potential mismatches between trial-specific estimands: in endpoint definitions, treatment implementations, study target populations, summary effect measures or intercurrent event strategies

There is scarce evidence that even relatively common intercurrent events, e.g. treatment switching, are reported in evidence syntheses or accounted for, particularly in analyses involving older studies

Cochrane has not embraced the estimands framework and explicitly advocates for ITT (“treatment policy”)

# Estimands: opportunities in HTA

Alignment between decision-making contexts is challenging…different policy questions require different research questions, decision problems, estimands

Nevertheless, greater recognition of the estimands framework is beneficial to:

- Harmonize regulatory and HTA language; valuable for convergence across guidance documents and parallel scientific advice and consultation
- Pose HTA-specific research questions earlier in drug development (e.g., at the trial design stage or earlier), so that needs can be met at the time of HTA decision-making
- Identify sources of uncertainty in HTA decision-making and of misalignment in evidence syntheses
- Define HTA research questions more clearly and less ambiguously
- Improve the practice of evidence syntheses

# Acknowledgements

I gratefully acknowledge Antonia Morga, Pepa Polavieja, Anders Gorst-Rasmussen and the EFSPI/EFPIA/PSI late phase HTA and RWE estimands Implementation Working Group. This presentation was strongly motivated by their insightful comments.

| | |
|---|---|
| Arthur Allignol | Daiichi-Sankyo |
| Amel Besseghir | ClinChoice |
| Khadija Rantell | Medicines & Healthcare products Regulatory Agency |
| Frank Kleinjung | BAYER |
| Antonio Remiro-Azocar | |
| Lotte Husemoen | novo nordisk® |
| Pepa Polavieja | |
| Anders Gorst-Rasmussen | |
| Antonia Morga | astellas |
| Barbara Rosettani | Bristol Myers Squibb™ |