Ensuring Estimand and Population Consistency Across Health Technology Assessment

Sylvaine Barbier [1], Chloë Hardern [2], Catriona Crossan [3], Mateusz Nikodem [4]
[1] Inizio Ignite Putnam, France, [2] Inizio Ignite Putnam, UK, [3] Inizio Ignite Putnam, Ireland, [4] Inizio Ignite Putnam, Poland

HTA39

Introduction

Health technology assessment (HTA) submissions integrate multiple interdependent evidence components, including:

- Systematic Literature Reviews (SLRs),
- Indirect Treatment Comparisons (ITCs),
- Cost-Effectiveness (CE) models, and
- Global Value Dossiers (GVDs).

Failure to properly link these components, particularly regarding populations, outcomes definitions and estimands, may compromise internal validity and lead to misleading clinical and economic conclusions. This study aims to develop a practical guide to support internal consistency.

Objective & Methods

- A structured methodological assessment was conducted to identify critical "linking points" where misalignment commonly occurs.
- They were mapped across the evidence generation, with emphasis on alignment of target populations, estimands (including marginal versus conditional estimands), outcome definitions, treatment strategies, and time horizons.
- Identification of linking points was informed by HTA methodological guidance and expert review.

Results

Seven key linking points requiring alignment were identified (see **Table 1**).

A diagram of how these points link across various evidence components is presented in **Figure 1**.

Failure to address these linking points may bias estimates of clinical and economic value. Therefore, in **Table 1**, we present examples of each linking point, explain what can go wrong, and provide guidance for how to mitigate issues.

We present a Checklist in **Figure 2** to summarize the guidance and aid the development of a consistent evidence generation package.

Figure 1

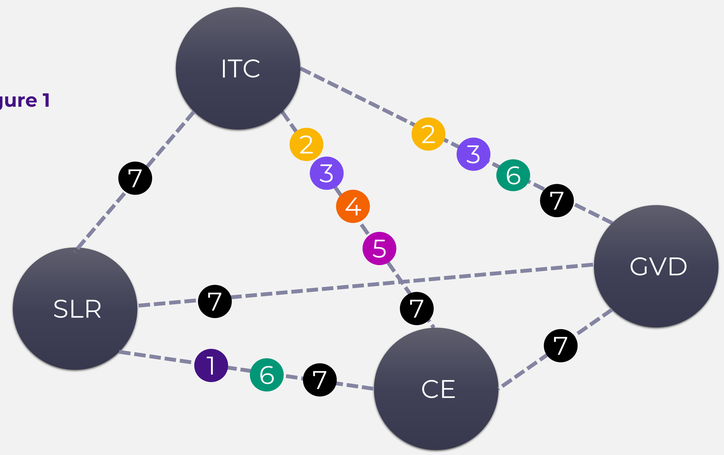

Figure 2

Checklist

SLR

- ☑ Does the specified population match the decision problem?
- ☑ Are the outcomes defined consistently?
- ☑ Are the effect modifiers identified?

ITC

- ☑ Is it consistent with the SLR and interpretable in context of the decision problem?
- ☑ Is it methodologically valid & adapted to the country-specific requirements?

CEM

- ☑ Is it aligned with the SLR and ITC?

GVD

- ☑ Is it strictly based on the findings referenced in the SLR/ITC/CEM?
- ☑ Is the content aligned with the narrative relevant for this HTA submission?

Table 1: Summary of linking points, what can go wrong, and guidance

	Linking point	Example	What can go wrong?	Guidance
1	**Consistency between SLR PICOS criteria, populations parameters and outcomes used in CE models**	The **SLR eligibility criteria** capture the population, comparators, follow-up periods and outcomes needed to populate the model, such as efficacy, utility, costs and long-term outcomes	• The SLR may be designed around the clinical evidence network but **fail to capture studies** needed for model inputs such as baseline risks, utilities, discontinuation, adverse events or long-term outcomes • The SLR may extract outcomes in a way that does **not match model** health states, cycle length or target population • This can lead to **late structural modelling rework**, **evidence gaps** or **poorly justified assumptions**	• Agree on the **model decision problem** before finalising the **SLR protocol** • Map each **model input** to the relevant PICOS elements and extraction requirements • Distinguish "essential" outcomes from exploratory outcomes and review alignment at protocol stage
2	**Alignment of ITC estimands with target decision population, including appropriate choice between marginal and conditional estimands**	In different trials, treatment effects may be estimated using models that yield **conditional (covariate-adjusted)** or **marginal (population-average) effects**. These may differ, particularly for non-collapsible measures (e.g. odds ratios, hazard ratios).	• Mixing conditional and marginal effects across trials can **bias ITC results**. • Even when aligned, using an estimand that does not match the decision context may limit **applicability to the target population**. • This can lead to **inconsistencies** in cost-effectiveness analyses.	• Clearly **define the target estimand** in line with the decision problem (typically population-average effects for HTA). • Ensure that ITC methods estimate effects aligned with this estimand. • Where possible, obtain marginal (population-level) effects by applying **conditional models** to the target population (e.g. via prediction and averaging) [1]
3	**Assessment of heterogeneity across trials populations included in the ITC and their transportability to the target population**	Trials included in the ITC may differ in **patient characteristics** (e.g. comorbidities, disease severity, treatment history). These populations may not reflect the target population for decision-making.	• **Differences in patient characteristics** across trials may not be fully accounted for, leading to biased relative treatment effects. • Results may not be **generalizable** to the target population if trial populations are not representative. • This can lead to **biased estimates** of comparative clinical efficacy and cost-effectiveness.	• Assess **heterogeneity** in patient characteristics across trials and evaluate their relevance to the target population. • Where needed, **adjust for cross-trial differences** (e.g. using population-adjusted methods) • Ensure that relative effects are estimated for, or appropriately transported to, the **target population**
4	**Coherence between relative treatment effects from ITCs and absolute event rates used in CE models**	ITC-derived odds ratios, hazard ratios or rate ratios are applied to baseline risks to estimate **absolute probabilities** of clinical outcomes such as response, progression, discontinuation or adverse events in the economic model	• **Relative effects** may be applied to inappropriate baseline risks, different endpoint definitions or incompatible time points • This can produce **biased absolute risks** or **implausible probabilities**	• **Pre-specify** how relative effects will be converted into absolute model inputs • Ensure the baseline risk source, time horizon and endpoint definition are **consistent** with the ITC • Validate resulting absolute risks against **observed trial data** and clinical expectations
5	**Appropriate estimation of baseline risks**	Event risk may be **non-linearly (e.g. exponentially) related** to continuous risk factors such as age. Risk may appear very low for most patients below a certain threshold but becomes substantially higher in those above it	• **Population-level event risk** and associated costs and resource use may be **underestimated** when high-risk subgroups are not properly captured • Treatment comparisons may be biased if **high-risk subgroups** are not adequately represented, as they can disproportionately drive overall results.	• Estimate risk at the individual level by **simulating the distribution of key patient characteristics** (e.g. age) in the target population • **Aggregate these predictions** to obtain population-level estimates
6	**Alignment of treatment sequences and handling of treatment switching across evidence synthesis and CE models**	Trials may include **crossover**, **rescue medication** or **switching** after non-response, while the economic model may represent first-line and subsequent-line **treatment sequences**	• ITC estimates may reflect trial-specific switching or rescue therapy patterns that are **inconsistent with the treatment sequence modelled** in the CEM • Post-discontinuation treatments, retreatment effects or subsequent-line efficacy may be **ignored** in the economic model	• **Define the treatment pathway** and modelled sequence early • Align ITC handling of crossover, rescue therapy and discontinuation with **model assumptions** • Specify evidence needs for subsequent treatments and **test uncertain sequence assumptions** in scenario analyses
7	**Consistency of endpoint and subgroup definitions across analyses**	Outcomes such as response and remission, and subgroups such as treatment-experienced patients, are **defined consistently** across the SLR, ITC and model	• **Similar endpoint names** may mask different definitions, thresholds, assessment timings or censoring rules • **Subgroups** may not be comparable across trials, causing inappropriate pooling or model inputs that do not reflect the intended population	• Create a **shared definitions log** for endpoints and subgroups • Check **comparability** before pooling or applying estimates in the model • Where definitions differ, justify harmonization strongly and explore sensitivity analyses

Discussion & Conclusion

- Internal consistency is critical for generating valid, decision-relevant evidence.
- Misalignment of populations, outcomes, estimands, and treatments may introduce bias in HTA conclusions, depending on the nature and direction of the mismatch.
- The proposed guide provides a structured framework to improve coherence and credibility of HTA submissions.

References
[1] Phillippo DM, Remiro-Azócar A, Heath A, Baio G, Dias S, Ades AE, Welton NJ. Effect modification and non-collapsibility together may lead to conflicting treatment decisions: A review of marginal and conditional estimands and recommendations for decision-making. Res Synth Methods. 2025 Mar;16(2):323-349. doi: 10.1017/rsm.2025.2. Epub 2025 Mar 10. PMID: 41626969; PMCID: PMC12527544.

Abbreviations
CEM, cost-effectiveness model; GVD, global value dossier; HTA, health technology assessment; ITC, indirect treatment comparison; SLR, systematic literature review

CONTACT
Sylvaine Barbier
Sylvaine.Barbier@putassoc.com

Find out more at Inizioignite.com