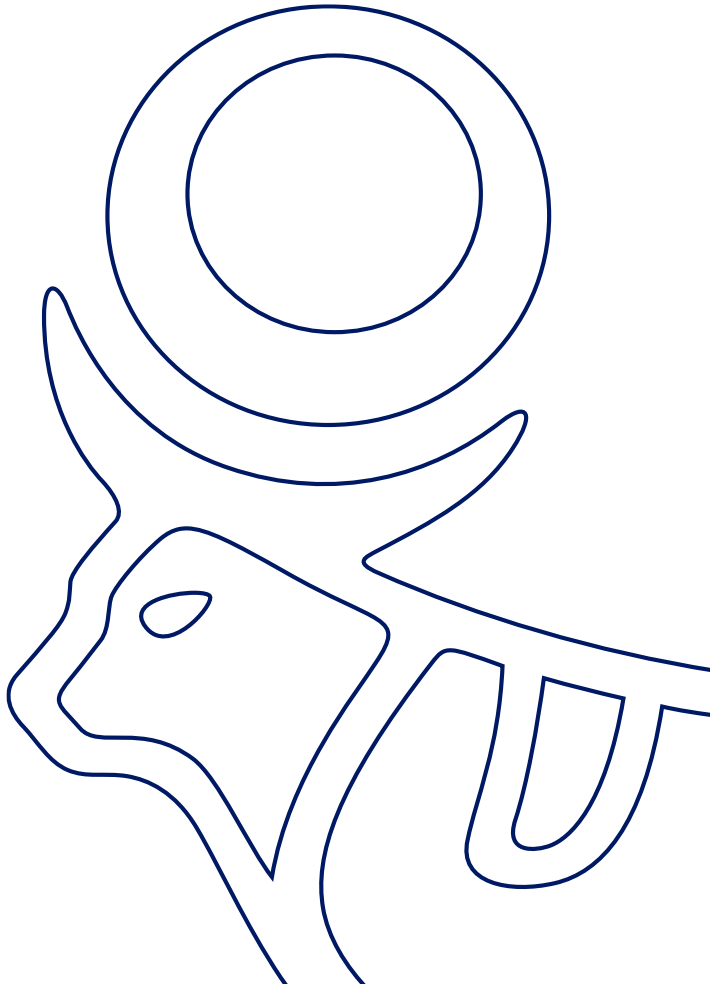
UNORDINARY DRIVES CHANGE

novo nordisk®

Is Transferring Evidence Across Europe a Herculean Task?

Challenges of Transportability Analyses in the Context of European Union Joint Clinical Assessments



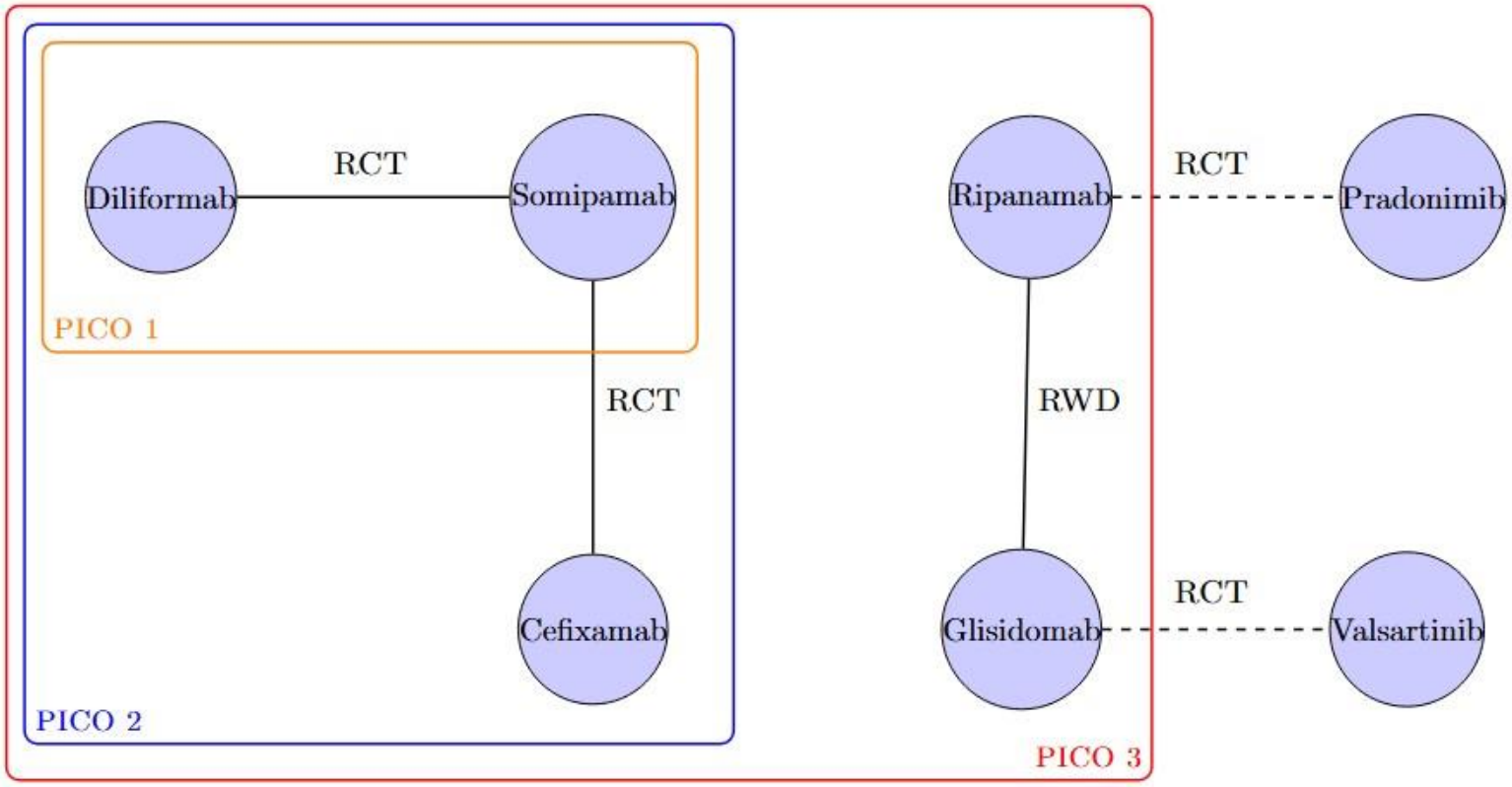
Antonio Remiro-Azócar, PhD

Biostatistics Methods, Medical and Translational Science, Novo Nordisk

ISPOR Europe, Glasgow, 10th November 2025

Inventiva Therapeutics is preparing for EU JCA…

The manufacturer of Diliformab, Inventiva Therapeutics, faces an increasingly complex and fragmented targeted oncology landscape when planning to submit its EU Joint Clinical Assessment dossier…
Clinical practice is heterogeneous and there are several similarly positioned comparators across member states



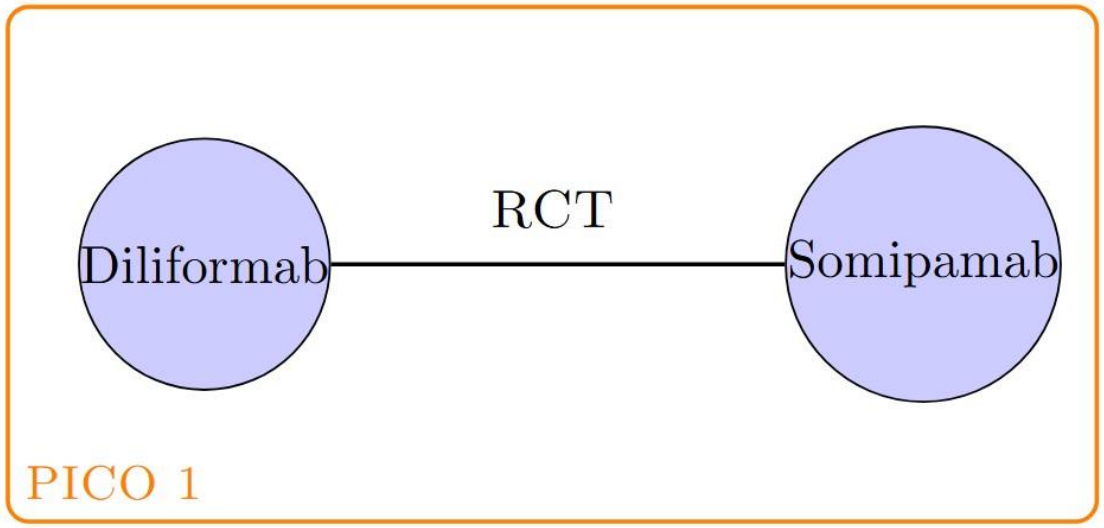
- Pivotal active-controlled Phase 3 RCT versus Somipamab
- Somipamab is licensed and treating a notable proportion of EU patients, based on robust clinical RCT evidence versus Cefixamab
- However, Cefixamab is still reimbursed as standard-of-care in a few countries
- Moreover, “off-label” Ripanamab and Glisidomab still feature in some clinical practice guidelines based on historical RCT results

PICO 1

Inventiva Therapeutics has conducted a Phase 3 RCT comparing Diliformab to Somipamab, the standard-of-care in most European countries

The RCT is a large, double-blind, trial with careful randomization, pre-specification and multiplicity control, and very little missingness and measurement error

Regulatory approval from EMA is eagerly anticipated



However, when initiating preparations for the EU JCA submission, Inventiva's HEOR team takes notice of an issue…

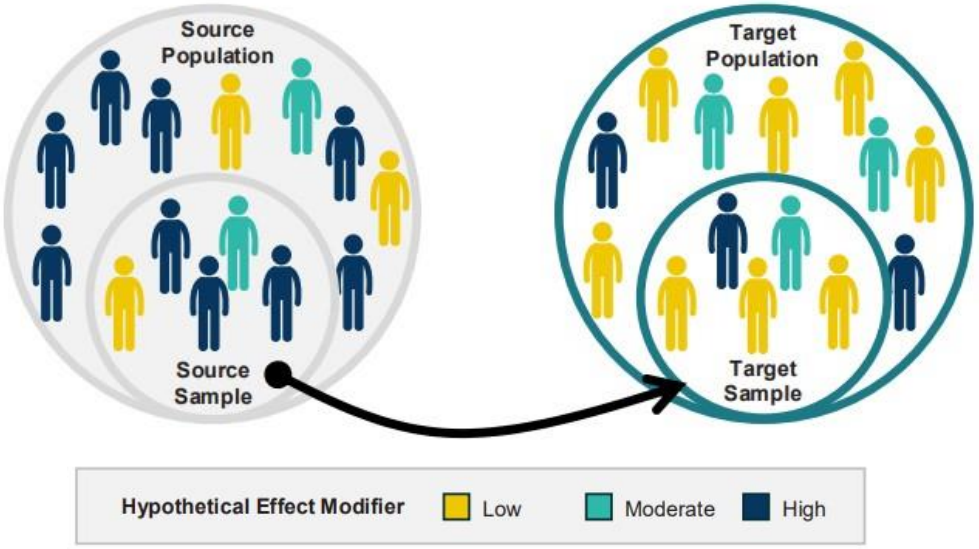
Due to suboptimal engagement with stakeholders and the failure to secure local and European scientific consultation slots, there is a mismatch between the population of the pivotal trial and that of PICO 1

EU JCA domains of certainty for the RCT: high internal validity, high precision, low external validity

A proposal from Scientifico Analytics…

A vendor, Scientifico Analytics, offers a methodological solution to the health technology developer's problem…

TRANSPORTABILITY ANALYSIS!



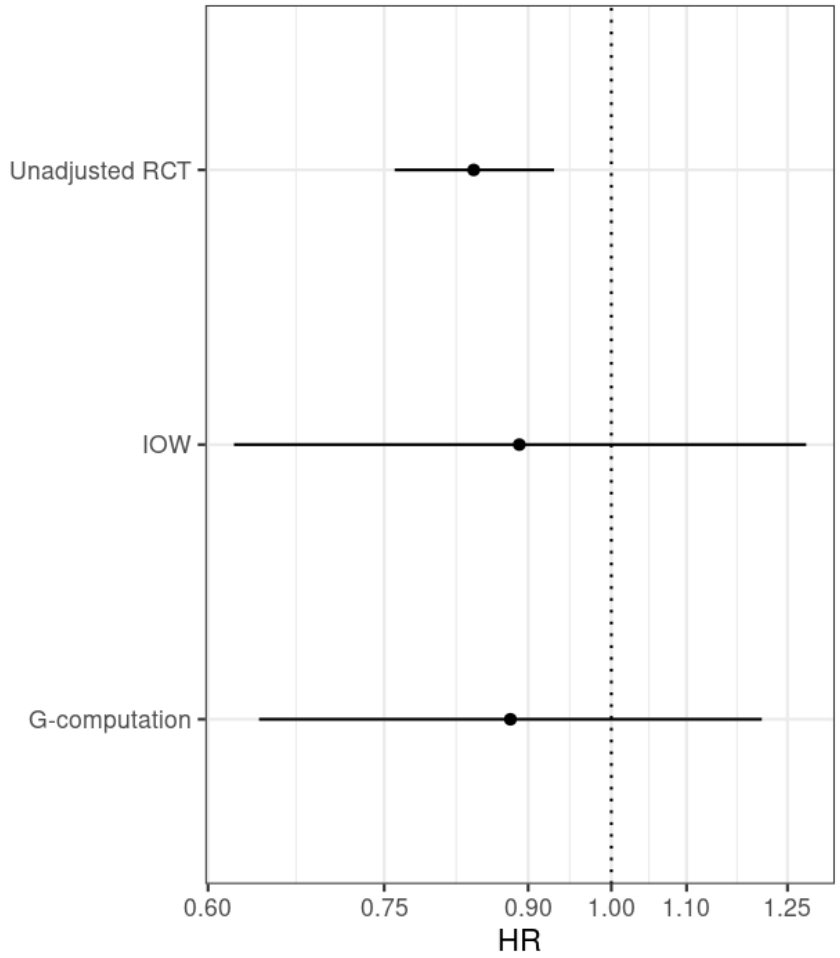
Vuong et al, 2025. Systematic review of applied transportability and generalizability analyses: A landscape analysis. Annals of Epidemiology

External validity ⬆ ⬆ ⬆ ⬆ ⬆

- "Transporting" the results of the pivotal RCT to an external European registry, which includes a sample of patients that is representative of the PICO 1 target population
- The vendor argues that the transportability analysis will improve external validity and help meet the evidentiary requirements of PICO 1
- The proposed "base case" methodology is an Inverse Odds of trial participation Weighting (IOW) approach that weights the pivotal RCT so that key "effect modifiers" are balanced with respect to the target sample
- A G-computation procedure that models and predicts outcomes, conditional on effect modifiers, is also considered

PICO 1 results

- The "naïve" RCT analysis produces a hazard ratio (HR) of 0.84 (95% CI: 0.76 – 0.93), statistically significant at the 5% level
- The IOW analysis produces HR = 0.89 (95% CI: 0.62 – 1.28), reducing the original sample size from 623 to an effective sample size of 126 after weighting
- The G-computation analysis gives HR = 0.89 (95% CI: 0.64 – 1.21)
- Both transportability analyses shift the point estimate towards the null (HR=1) and considerably inflate the interval width:
  - IOW: variance increases due to extreme weights
  - G-computation: variance increases due to extrapolating an outcome model with treatment-covariate interactions



Unadjusted analysis: high internal validity, high precision, low external validity

Transportability analyses: high internal validity, low precision, high external validity

PICO 1 predicament: which is preferrable?

Unadjusted analysis: high internal validity, high precision, low external validity

Transportability analyses: high internal validity, low precision, high external validity

In my opinion, the transportability analysis is preferrable:

- The unadjusted analysis is **over-precise** in the PICO 1 target population
- High precision is futile if it has been quantified inaccurately…but this is not acknowledged by the EU JCA guidelines
- The transportability analyses (particularly IOW) capture the uncertainty of estimation in an external target
- All analyses have high internal validity, which according to the JCA "Validity of Clinical Studies" guidance, is the "prioritary" dimension of certainty (*this is a value judgment*)

Nevertheless, there are reasons to suggest that any assessor purely following the JCA guidance might place the first over the second:

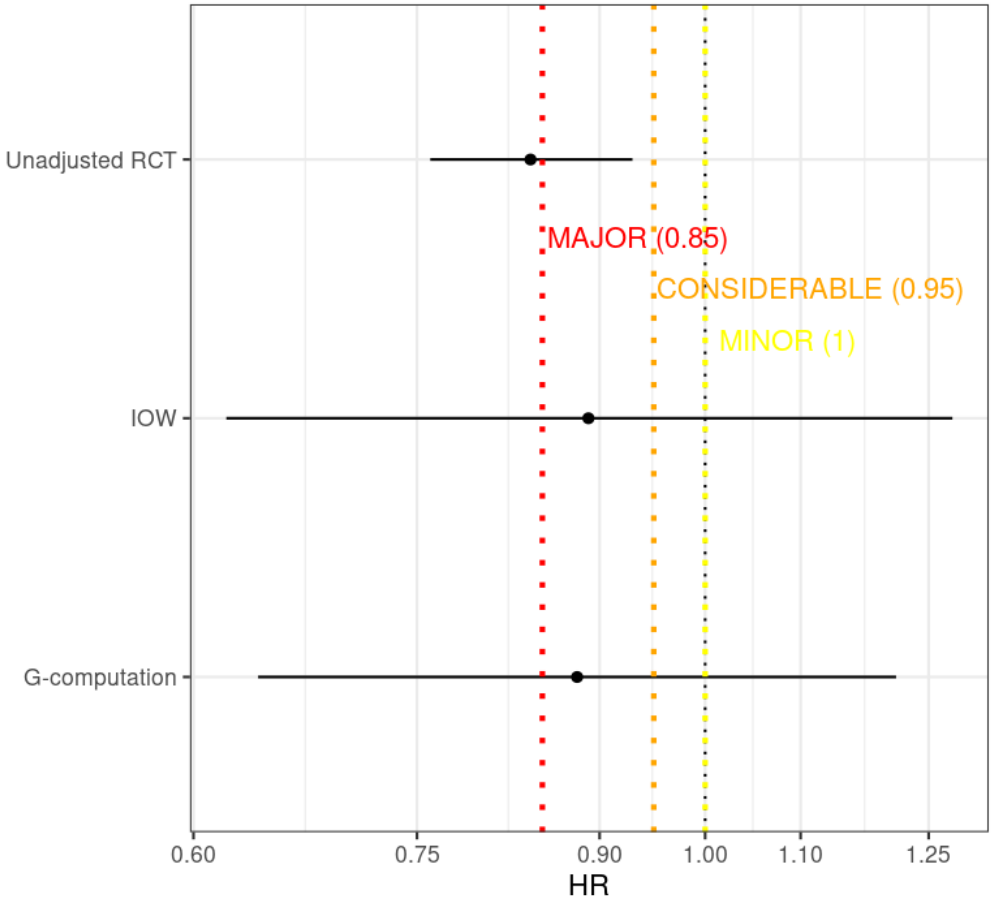
- The assessment of external validity is a value judgment left to the member states
- Any changes to the pre-specified trial statistical analysis plan to accommodate mismatches with the requested PICO will be flagged as "post-hoc", devaluing the perceived strength of evidence
- Moreover, there is a potential "shifted null hypothesis testing" requirement

Shifted null hypothesis testing (SNHT)

According to EU JCA methodological guidelines, population-adjusted estimates must be tested against a threshold, shifted away from the null by a “large enough” magnitude, due to increased “researcher degrees of freedom” and uncertainty due to “missing effect modifiers”

- Who sets the threshold? The health technology developer or the individual member states? The guidance is contradictory
- If the HTD sets the threshold, this implies a value judgment about the uncertainty that member states are willing to accept
- If the member states set the threshold, it cannot be pre-specified, which renders the shifted null hypothesis test invalid
- How is the threshold determined? (*value judgment*) There is no consensus for any outcome or therapeutic area
- The tests are severely underpowered and prone to Type 2 error

Example: the German “extent of added benefit” thresholds



PICO 1: conclusions

In this example, the transportability analyses would always fail to reject the shifted null hypothesis tests because the SNHT threshold undercuts the upper bound of the confidence interval

Unadjusted analysis: **high internal validity**, **high precision**, **low external validity**, **pre-specified**, **some SNHTs rejected**

Transportability analyses: **high internal validity**, **low precision**, **high external validity**, **post-hoc**, **failure to reject SNHTs**

There are very valid reasons to perform the transportability analyses…

…but the health technology developer might have fared better with the unadjusted analysis (and a qualitative assessment of external validity)

The shifted null hypothesis testing requirement should be removed from EU JCA guidance

There are less crude and better methods to offset researcher degrees of freedom, which directly cater to the needs of HTA decision-makers, such as quantitative bias analysis

PICO 2

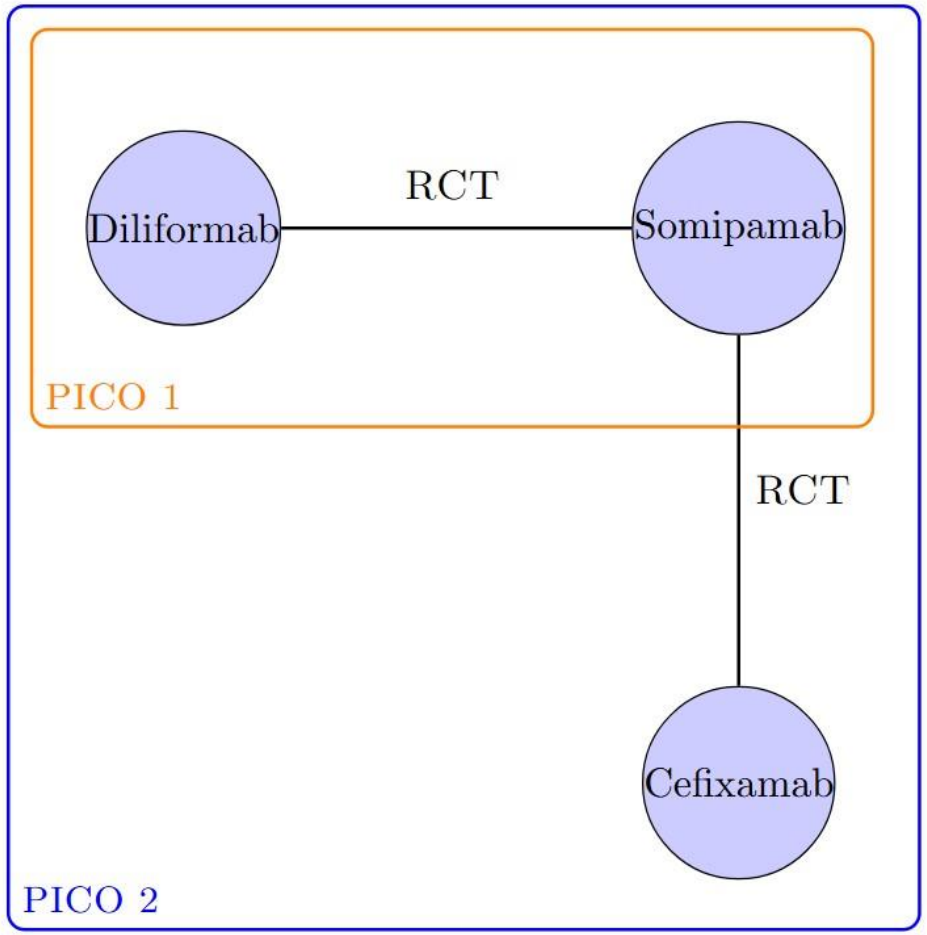
There is no head-to-head RCT between Diliformab and Cefixamab

However, Panacea Pharmaceuticals previously conducted its own “pivotal” Phase 3 RCT of Somipamab versus Cefixamab to support the regulatory approval of Somipamab

The RCT was a large, double-blind, trial with careful randomization, pre-specification and multiplicity control, and very little missingness

Panacea Pharmaceuticals was heavily engaged in scientific consultations at the time, and carefully selected its RCT population so that it closely matched the patients in European clinical practice

To compare Diliformab versus Cefixamab and meet the evidentiary requirements of PICO 2, Inventiva Therapeutics is set to perform an indirect treatment comparison, anchored via Somipamab



Another proposal from Scientifico Analytics…

The vendor, Scientifico Analytics, notices that the distribution of effect modifiers is different between the Phase 3 RCT sponsored by Inventiva Therapeutics and that sponsored by the “competitor” Panacea Pharmaceuticals

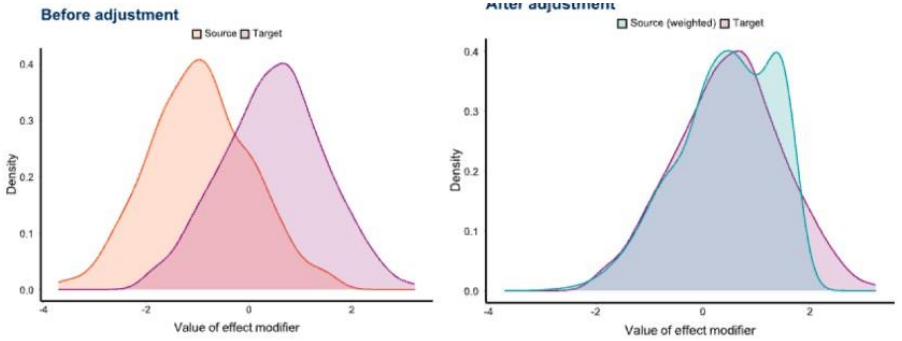
Awkwardly, the vendor also notes that the “competitor” RCT population is more representative of the PICO 2 population

The vendor suggests avoiding unadjusted anchored indirect comparisons (Bucher) for valid reasons:

- **Poor internal validity** (at the “meta” level): bias if there is treatment effect heterogeneity over baseline characteristics that vary in distribution over studies (and for the hazard ratio, these can be “purely prognostic”)
- **Poor external validity**: unexplained heterogeneity, estimates are not produced in any specific target population
- **Over-precision**: variance underestimation due to ignoring cross-trial differences in baseline characteristics

The vendor proposes several methodological solutions to overcome these challenges:

MATCHING-ADJUSTED INDIRECT COMPARISON (MAIC)
SIMULATED TREATMENT COMPARISON (STC)
MULTILEVEL NETWORK META-REGRESSION (ML-NMR)



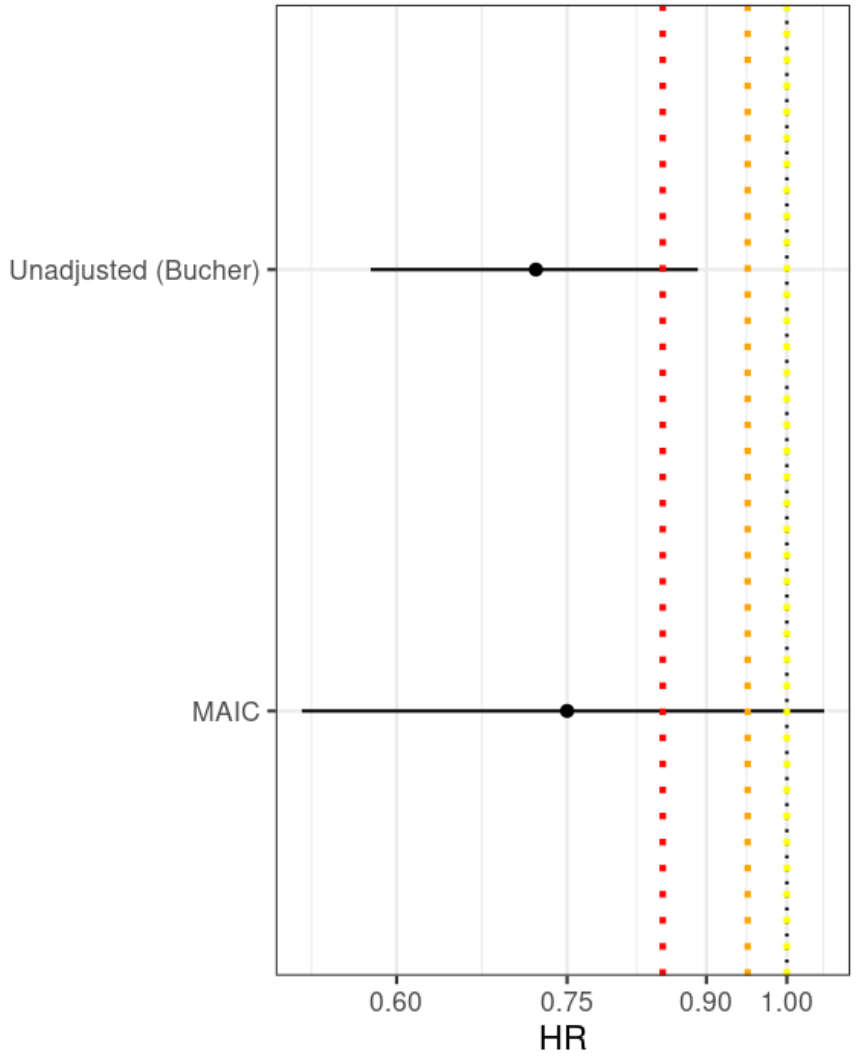
Vuong et al, 2025. Systematic review of applied transportability and generalizability analyses: A landscape analysis. Annals of Epidemiology

PICO 2: results

- The unadjusted “naïve” analyses of the RCTs produce HR=0.84 (95% CI: 0.76 – 0.93) for Diliformab versus Somipamab and HR=0.86 (95% CI: 0.77 – 0.96) for Somipamab versus Cefixamab

- Unadjusted anchored indirect comparison (Bucher):
  HR for Diliformab versus Cefixamab = 0.84 x 0.86 = 0.72
  Standard error (in log HR scale) ≈ 0.11
  HR (95% CI) for Diliformab versus Cefixamab = 0.72 (0.58 – 0.89)

- Matching-adjusted indirect comparison (MAIC)
  HR for Diliformab versus Somipamab = 0.87 (95% CI: 0.63 – 1.20)
  HR for Diliformab versus Cefixamab = 0.87 x 0.86 = 0.75
  Standard error (in log HR scale) ≈ 0.17
  HR (95% CI) for Diliformab versus Cefixamab = 0.75 (0.53 – 1.05)



Unadjusted Bucher analysis: low internal validity (“meta” level), high precision, low external validity

Population-adjusted MAIC analysis: high internal validity (“meta” level), low precision, high external validity

PICO 2: conclusions

Again, there are very principled reasons to perform the population-adjusted analysis, but the health technology developer might have fared better with the Bucher analysis (and a qualitative assessment of exchangeability)

Unadjusted Bucher analysis: low internal validity, high precision, low external validity, some SNHTs rejected

Population-adjusted MAIC analysis: high internal validity, low precision, high external validity, failure to reject SNHTs

The JCA methodological guidelines implicitly impose a fourth domain of uncertainty, **methodological complexity or researcher degrees of freedom (RDOF)**:

- If additional RDOF are clearly pre-specified and justified, alongside sensitivity analysis, why is SNHT required?
- Why is SNHT required for MAIC and not for Bucher? Both are threatened by unmeasured effect modifiers
- Member states can arbitrarily increase the complexity of the evidence network by adding superfluous (sub)populations and comparators. Why is the increase in methodological complexity penalized if it is warranted to increase validity?
- Negative wording about population adjustment: it has “to be applied with the utmost care” , is only useful to “confirm the results of (unadjusted) network meta-analysis”, and “is often more suitable as an exploratory analysis”

PICO 3

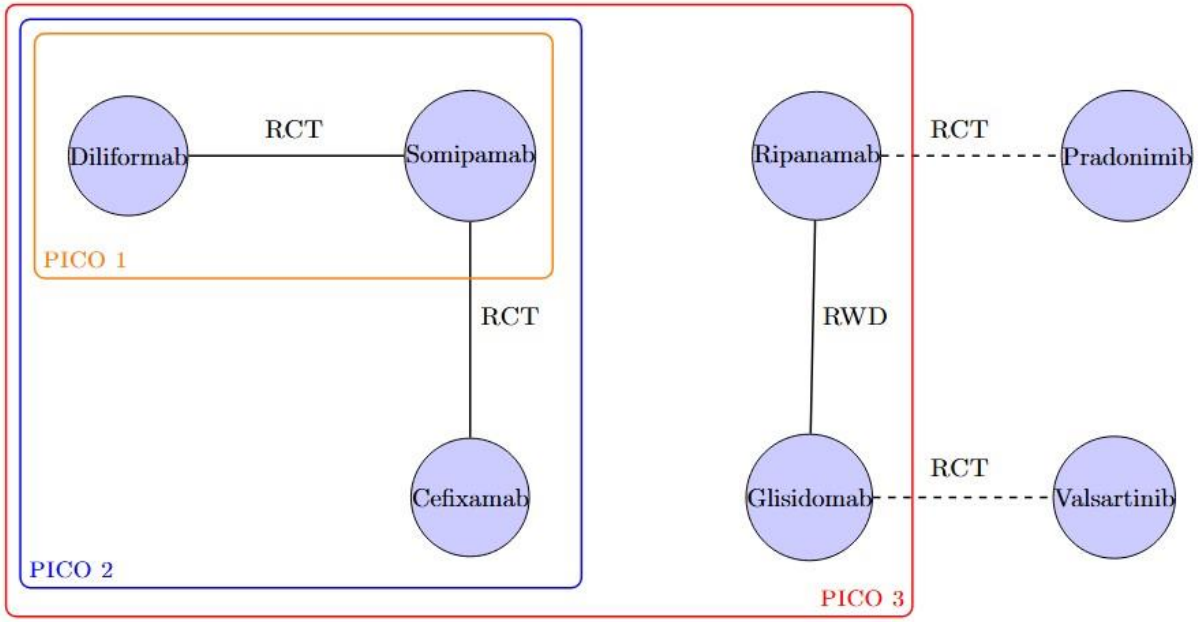
Inventiva Therapeutics has access to European subject-level RWD for patients under Ripanamab and Glisidomab

An unanchored (externally controlled) indirect comparison for Diliformab versus these treatments can be performed in the registry population

However, the HTD ultimately decides not to do so:

- Very negative wording about unanchored comparisons in the EU JCA methodological guidelines
  - “Highly problematic”
  - “Insufficient”
  - “Unlikely to provide a meaningful estimate”

- The SNHT thresholds likely set the evidentiary bar too high, as appropriate propensity score-based methods for confounding adjustment will (justifiably) increase the variance



Disconnected networks are inevitable as a by-product of a highly inclusive scoping process that aims to meet the needs of all 27 EU member states!

Transportability: a “herculean” task in EU JCA

The less principled “unadjusted” analyses:

- Diliformab versus Somipamab: HR = 0.84 (95% CI: 0.76 – 0.93) “considerable added benefit”*
- Diliformab versus Cefixamab: HR = 0.72 (95% CI: 0.58 – 0.89) “considerable added benefit”*

*Note: value judgment which would have been made, following the JCA process, at the member state level using the German thresholds

The more principled transportability/population-adjusted analyses:

- Diliformab versus Somipamab: HR = 0.89 (95% CI: 0.62 – 1.28) “no added benefit”*
- Diliformab versus Cefixamab: HR = 0.75 (95% CI: 0.53 – 1.05) “no added benefit” * !!!!!!!!!!

In both cases, the vendor has followed the optimal approach to enhance validity with respect to the specific research/policy questions (PICOs)…although admittedly the HTD could have better designed its pivotal RCT

But shifted null hypothesis tests depend on precision…and transportability analyses (justifiably) decrease precision

The current JCA guidelines, regrettably, do not incentivize the development of **quantitative** methodology for transportability/population adjustment…the HTD might have fared better with a qualitative assessment:
“Given that lack of external validity as compared with internal validity is usually more straightforward to detect, it might be sufficient to assess any issues (…) on a case-by-case basis using qualitative descriptive methods“ (Validity of Clinical Studies, p. 10)

Some final remarks on EU JCA

In our example (PICO 2 and PICO 3), validity and certainty were unduly devalued by PICOs from a minority of states, arguably redundant for most countries

PICO selection is permeated by implicit *value judgments* in the definition of (sub)populations and comparators

There is no EU JCA guidance available so far on evidence-based methods for PICO selection and prioritization

Any member state can arbitrarily increase the complexity and heterogeneity of an evidence network through potentially superfluous (sub)populations and comparators, **impacting the perceived validity, certainty and strength of evidence**…

…but attempts to compensate for this analytically via increased “researcher degrees of freedom” are penalized!