Considerations for Methodological Innovation for Indirect Treatment Comparisons in Pan-European HTA

Antonio Remiro-Azócar, PhD
Methods & Outreach, Novo Nordisk
9th EFSPI Regulatory Statistics Workshop, Basel
13th September 2024

Agenda

1. **Background**
   Covariate adjustment is necessary for indirect treatment comparisons (ITCs) in EU HTA

2. **Perception of covariate-adjusted ITCs**
   The implementation and conduct of covariate-adjusted ITCs has generally been poor

3. **Fit-for-purpose methods**
   Limitations of existing methods and desirable properties for new methods

4. **Conclusion**
   Some suggestions for future steps

The views in this presentation are my own and do not necessarily represent those of Novo Nordisk

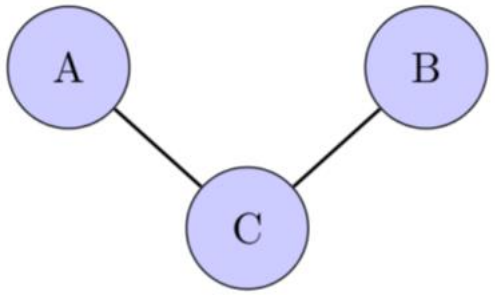
1. Background

Covariate adjustment is necessary for indirect treatment comparisons (ITCs) in EU HTA

Terminology

Indirect treatment comparisons (ITCs) can be **anchored** or **unanchored**

ANCHORED COMPARISON



UNANCHORED COMPARISON

ITCs can be **unadjusted** or **covariate-adjusted**

- **Unadjusted** ITCs do not explicitly adjust for cross-study differences in baseline covariates: Bucher method, traditional network meta-analysis, naive unanchored comparisons
- **Covariate-adjusted** ITCs explicitly adjust for cross-study differences in baseline covariates: matching-adjusted indirect comparison (MAIC), simulated treatment comparison (STC), G-computation, multilevel network meta-regression (ML-NMR), propensity score weighting,…

Why perform covariate adjustment?

Unadjusted ITCs…

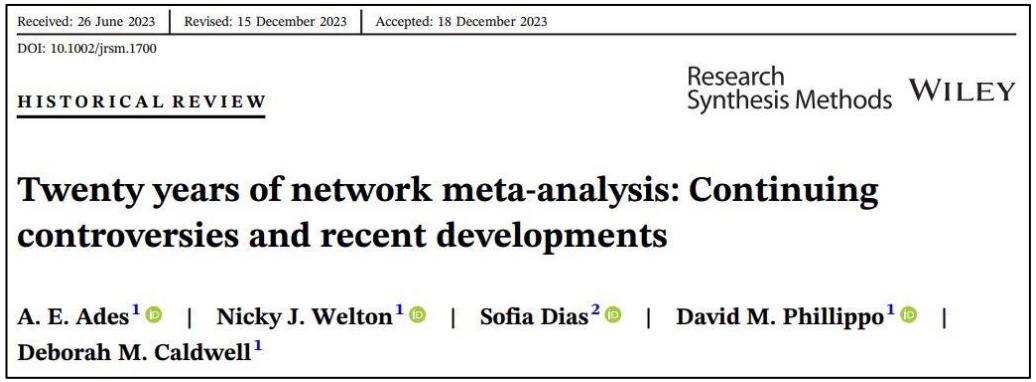
- Rely on a very strong assumption: unconditional exchangeability across studies
- Produce bias with cross-study imbalances in effect modifiers and/or prognostic variables
- Do not explain heterogeneity or explicitly produce estimates in any specific target population
- Ignore uncertainty due to cross-study differences in baseline covariates

Covariate-adjusted ITCs…

- Relax the exchangeability assumption by conditioning on baseline covariates
- Can reduce bias due to cross-study imbalances in effect modifiers and/or prognostic variables
- Explicitly produce estimates in specific target samples or populations
- Can account for uncertainty due to differences in baseline covariates across studies

Covariate adjustment is desirable for ITCs

The HTA context



Received: 26 June 2023 | Revised: 15 December 2023 | Accepted: 18 December 2023
DOI: 10.1002/jrsm.1700
HISTORICAL REVIEW
Research Synthesis Methods WILEY

Twenty years of network meta-analysis: Continuing controversies and recent developments

A. E. Ades[1] | Nicky J. Welton[1] | Sofia Dias[2] | David M. Phillippo[1] | Deborah M. Caldwell[1]

Received: 30 April 2024 | Revised: 24 May 2024 | Accepted: 8 July 2024
DOI: 10.1002/jrsm.1741
DISCUSSION
Research Synthesis Methods WILEY

Broad versus narrow research questions in evidence synthesis: A parallel to (and plea for) estimands

Antonio Remiro-Azócar[1] | Anders Gorst-Rasmussen[2]

- ITCs are used to inform HTA and reimbursement decisions
- HTA and reimbursement decisions are made for specific healthcare settings
- The research questions guiding these decisions should be similarly narrow, targeting specific treatment implementations, patient populations and outcomes
- Every effort should be made to reduce heterogeneity, so that treatment effect estimates are maximally relevant to policy decisions and target populations

The EU HTA Regulation

- The EU HTA Regulation introduces a centralized framework for the Joint Clinical Assessment (JCA) of new medicines
- The JCA scoping process is inclusive and aims to meet the diverse evidence needs of all 27 EU member states simultaneously
- The assessment scope (PICO) is determined by the EU states, prior to consolidation
- There is often limited consensus on **comparators** and **(sub) populations** across states, due to variation in clinical practice and reimbursement

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

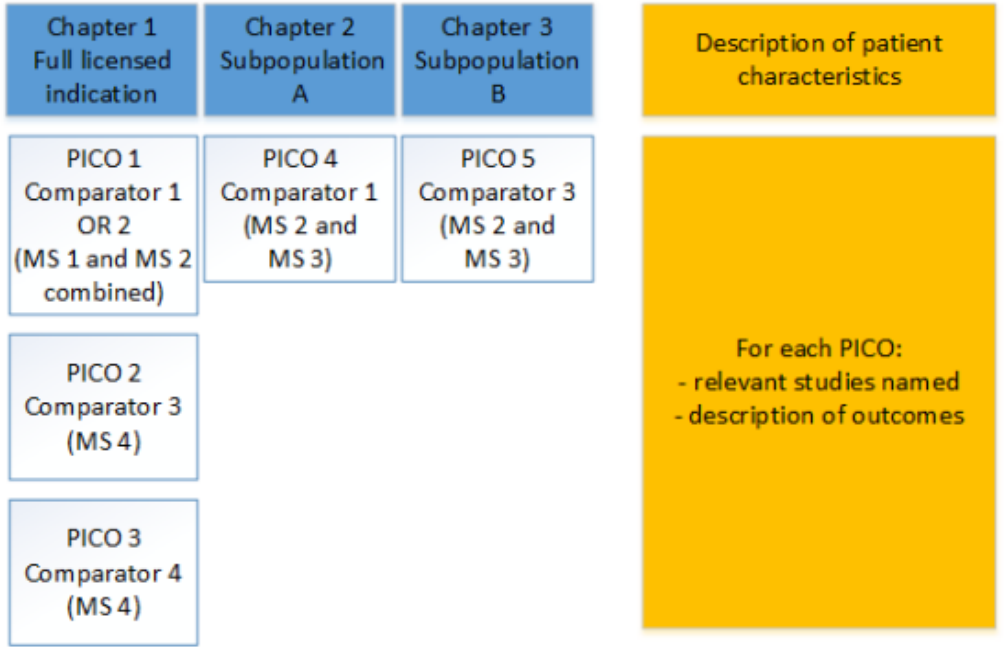


Figure 5-1: Data presentation according to PICO(s).

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

A multitude of PICOs

- van Engen et al (2024) follow EUnetHTA 21 Consortium guidance to determine the number of consolidated PICOs for two hypothetical products in two common oncology indications
- There is high variability in PICOs, **with many target (sub) populations**
- A substantial number of PICOs **require ITCs** due to unavailable head-to-head comparisons, some likely relying on small sample sizes from subgroup data and sparse networks

Table 3. Summary of PICO results

	1L NSCLC		3L MM	
	Base case (EU HTA reports)	**Base case + NICE report**	**Base case (EU HTA reports)**	**Base case + NICE report**
Populations	EMA label + 8 subpopulations	EMA label + 10 subpopulations	EMA label + 4 subpopulations	EMA label + 6 subpopulations
Comparators	9	9	8	9
Outcomes per PICO	28	28	45	45
Number of PICOs (% requested by single country)	10 (50%)	14 (50%)	16 (75%)	18 (78%)
PICOs requiring ITC	5	8	11	12
Number of analyses requested (% indirect analyses)	280 (50%)	392 (57%)	720 (69%)	810 (67%)

Abbreviations: European Medicines Agency; EU: European Union; ITC: indirect treatment comparison; HTA: health technology assessment; MM: multiple myeloma; NICE: National Institute for Health and Care Excellence; NSCLC: non-small cell lung cancer; PICO: population, intervention, comparator, outcome.

van Engen et al (2024), The Impact of Additive PICOs in a European Joint Clinical Health Technology Assessment

Part 1 conclusions

- The use of ITCs for HTA decision-making is expected to grow dramatically
- The target population has a fundamental role in scoping decision problems in the EU HTA Regulation, in the PICO (population, intervention, comparator, outcome) framework
- Covariate adjustment is necessary to maximize relevance with respect to the target population for decision-making in each PICO
- Covariate adjustment is necessary for ITCs in EU HTA

2. Perception

The implementation and conduct of covariate-adjusted ITCs has generally been poor

Current perception in Europe

Current (pre-EU HTA Regulation) acceptance for covariate-adjusted ITCs in Europe is low:

- Macabeo et al (2023) analyze HTA reports for solid tumour oncology treatments between April 2018 – April 2021 in England, France, Germany, Italy and Spain
- The overall acceptance rate of ITCs was suboptimal, 36/120=30%, with a rejection rate of 25/120=21%
- Unadjusted ITCs are more commonly used than covariate-adjusted ITCs and have greater acceptance rates overall (27/76=36% versus 8/24=33%)
- This is in a therapeutic area where covariate adjustment is appealing – innovative treatments with single-arm trials, accelerated approval requests, small sample sizes and sparse networks

Criticisms of covariate-adjusted ITCs:

- Increased complexity and ”researcher degrees of freedom”
- Poor reporting standards, lack of transparency in implementation
- Publication bias

These issues undermine confidence in the underlying methodology

Researcher degrees of freedom

Covariate adjustment requires additional choices, some subjective, on:

- Covariate adjustment methodology
- Covariate selection
- Model selection
- Sample refinement to align patient eligibility/selection criteria
- Balancing means and/or higher-order moments (weighting)
- Covariate simulation (outcome modeling)
- Target population
- Variance estimation
- Variance reduction approaches (weighting), e.g. trimming/truncation cut-offs

Sensitivity analyses to assess the implications of these choices are often not implemented:

- As of February 13, 2023, only 85/162 (52.5%) of MAICs, STCs or ML-NMRs in peer-reviewed publications conducted some sort of sensitivity analysis (Truong et al 2023)

Poor reporting standards

Increased complexity and researcher degrees of freedom requires more transparent reporting

As of April 2, 2022, according to a review by Serret-Larmande et al (2024):

- The rationale for covariate selection is not explained in 36% (104/288) of peer-reviewed publications
- Many (127/288, 44%) do not discuss whether covariates are prognostic factors or effect modifiers

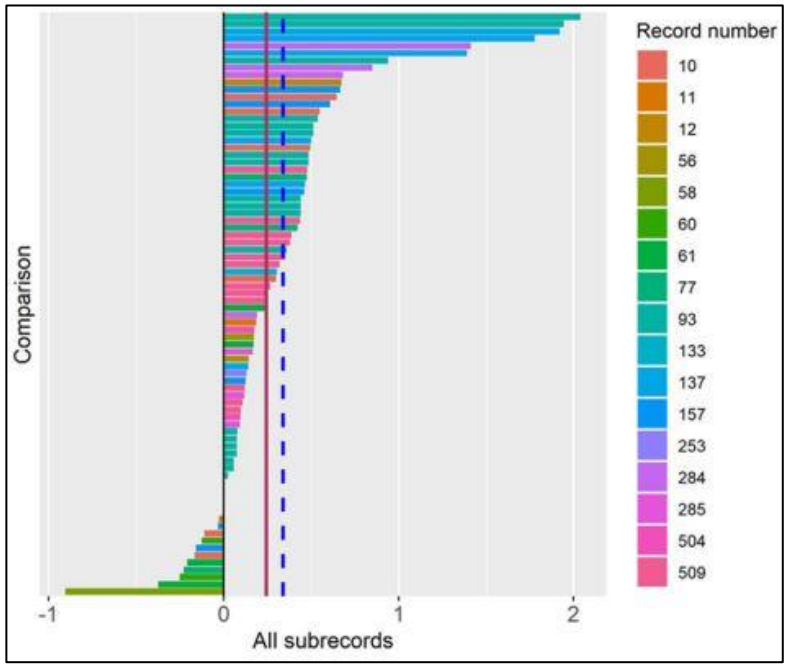
According to Truong et al (2023):

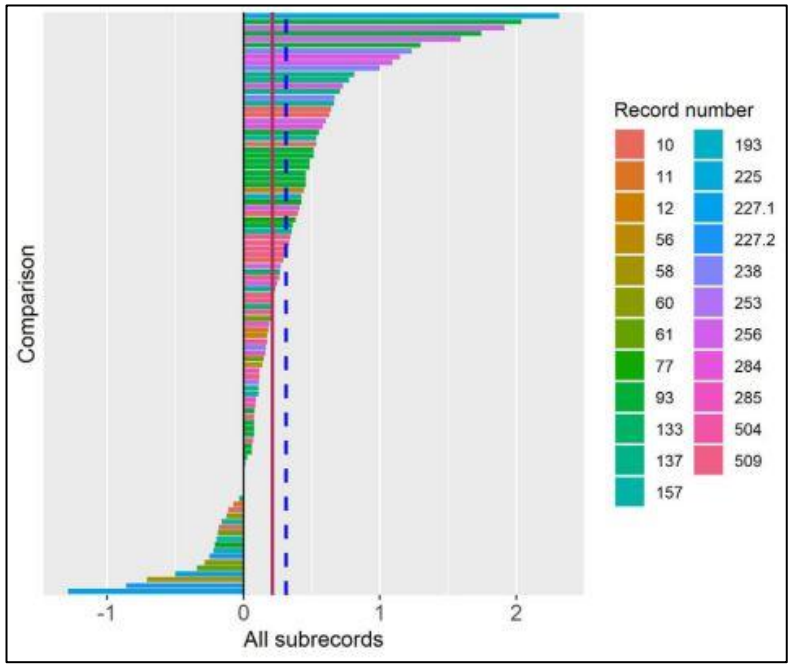
- Covariate selection is often based on the availability of covariates in all studies (67/162, 41%)
- Among studies using outcome modeling approaches (STC and/or ML-NMR), details about model selection and/or estimation, e.g. model-fitting diagnostics, are barely reported (3/18, 17%)

Publication bias

Cassidy et al (2023) empirically compare published treatment effect estimates from anchored MAICs with estimates from corresponding unadjusted anchored ITCs until February 25, 2019

MAICs present a more favorable point estimate for the ”index” treatment relative to the unadjusted approach





Cassidy et al (2023), A Comparison of Relative-Efficacy Estimate(s) Derived From Both Matching-Adjusted Indirect Comparisons and Standard Anchored Indirect Treatment Comparisons: A Review of Matching-Adjusted Indirect Comparisons

Poor implementation hurts methodological advancement

- Increased complexity and researcher degrees of freedom, lack of transparency and poor reporting increase uncertainty in the HTA decision-making process
- JCA guidelines suggest ”shifted hypothesis testing” – requiring a larger treatment effect to offset or penalize the additional uncertainty
- Assessors will prefer simpler, more familiar, methods if these are perceived to be more clear, transparent and convenient
- Similarly, asessors may prefer a more laissez-fare approach to heterogeneity (e.g. risk-of-bias tools), where potential biases/heterogeneity are documented and evidence downgraded, than sophisticated covariate adjustment methodology

Part 2 conclusions

- The implementation and reporting of covariate-adjusted ITCs are suboptimal in current practice
- Methodological implementation is not evolving as quickly as methodological developments…giving rise to a divergence between applications and development
- As the pace of methods development keeps accelerating, it is important to be aware of this "gap" (Jackson et al 2024)
- Recognizing the balance between methodological applications and innovations is key

3. Fit-for-purpose methods

Limitations of existing methods and desirable properties for new methods

Further progress required

- While improvements in implementation and reporting are necessary, further development and promotion of new methods is also required
- Covariate adjustment and causal inference methodologies continue to evolve quickly
- Further techniques will become available in the future context of ITCs
- Certain properties are desirable for reliable HTA decision-making
- Methods with these properties are ”better” and should be preferred

Currently available methodologies

OUTCOME MODELLING (STC, ML-NMR)

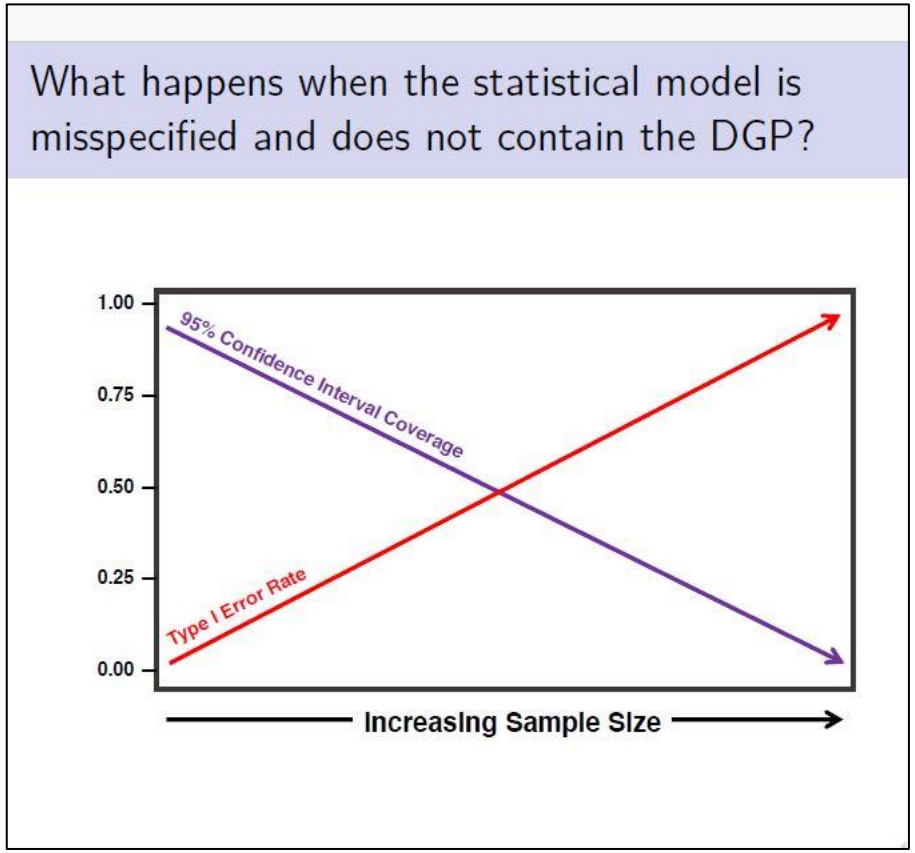
- Parametric model for the conditional outcome expectation given baseline covariates (and treatment)
- Relies on model-based extrapolation to improve statistical precision and efficiency with respect to weighting
- Susceptible to bias when extrapolating a mis-specified outcome model
- Model misspecification bias difficult to detect; an outcome model that seems approximately correct in the “index” study may not fit well in extrapolated regions
- Extrapolation uncertainty often not accounted for
- Can produce the treatment effect estimates that are required for HTA where there is limited overlap

ODDS WEIGHTING (MAIC)

- Parametric model for the conditional probability of trial assignment given baseline covariates
- Does not extrapolate; more “honest” uncertainty quantification
- MAIC is more “bias-robust” than than the standard “inverse weighting” modelling approaches
- Model misspecification bias easier to diagnose, MAIC (entropy balancing) directly enforces balance in covariate moments
- Extreme weights explicitly manifest high uncertainty
- Feasible weighting solutions may not exist where there is limited covariate overlap, e.g. convergence failures

Limitations of available methodologies

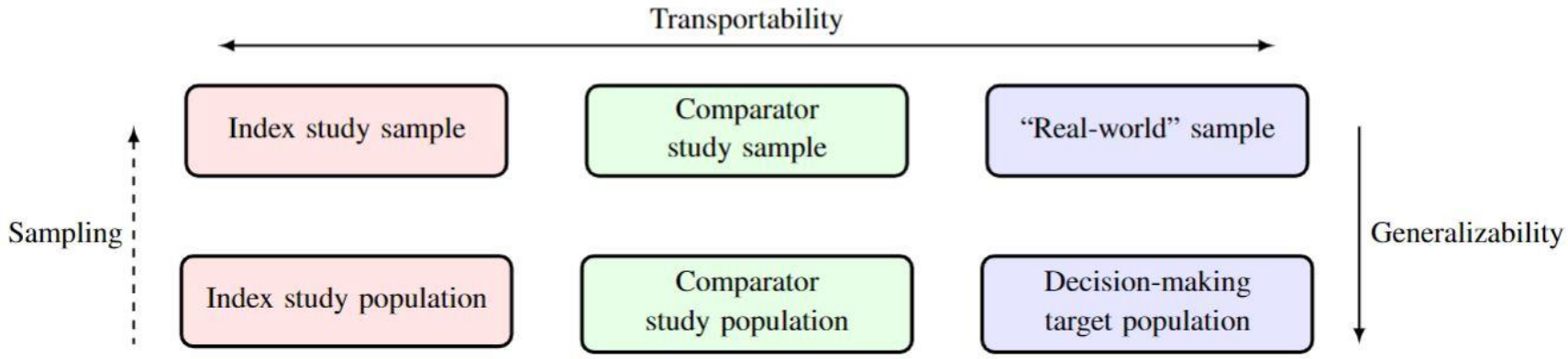
- Current approaches are singly robust (in most cases) and based on parametric modeling, relying on the correct specification of a single parametric model
- Strong parametric assumptions are often unsubstantiated and fail to reflect the complexities of the real data
- If the parametric model is incorrectly specified, the singly robust estimator is subject to bias and this bias does not decrease with sample size, at any rate
- As the sample size grows, the probability that interval estimates contain the target estimand shrinks to zero



van der Laan (2024), Targeted Learning for Causal Inference Using Real-World Data, ISPOR 2024

Limitations of MAIC, STC

- External validity with respect to the target population for HTA decision-making



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

- MAIC and STC are restricted to contrast treatments in the "comparator" study sample
- This may differ to the target population in the jurisdiction of interest

In anchored settings, ML-NMR can produce estimates in any specified target population…

- In any of the study samples included in the meta-analysis
- In an external source generated from real-world data, registries or observational studies

ML-NMR is clearly relevant to the EU JCA with many PICOs and target populations…

…*but is still a singly robust estimator relying on strong parametric assumptions*

The promise of doubly robust methods

- Doubly robust methods, such as augmented odds weighting and targeted maximum likelihood estimation (TMLE), may provide greater ”bias-robustness”
- These apply two working models: one for trial assignment conditional on covariates, and another for the outcome conditional on covariates (and treatment)
- Only one of the two models needs to be correctly specified to achieve unbiased estimation
- Doubly robust estimators should be less prone to model misspecification bias than singly robust estimators; they offer two opportunities for valid adjustment instead of one
- Nevertheless, they are still subject to bias where the two working models are parametric: both parametric models are likely incorrect!

The promise of data-adaptive estimators

More flexible machine learning (ML) estimators make fewer structural assumptions about the data generating-mechanisms and reduce the risk of model misspecification bias

Data-adaptive estimation within a doubly robust framework such as TMLE provides:

- Precision/efficiency…**while limiting unreasonable extrapolation with poor overlap**
- Good finite-sample performance…**by allowing for the use of slower-converging models**
- Valid statistical inference and uncertainty quantification…**by sample splitting (”cross-fitting”) to relax the Donsker condition, weaking the restrictions on the algorithms that can be used**

While ML may automate elements of model selection, it introduces other complexities and researcher degrees of freedom, with further challenges for the transparent reporting of evidence

ML: additional degrees of freedom

- Algorithm choice
  - What not to do: naively picking a default algorithm ”ad hoc”
  - **What to do: stack the candidate algorithms using an ensemble approach, e.g. Super Learner**

- Hyperparameter tuning/settings
  - What not to do: blindly use the default settings of the software package
  - **What to do: sensitivity analyses, outcome-blind simulations**

- Stochasticity, dependence on random number generator seed values
  - **What to do: sensitivity analyses, averaging strategies across seeds to stabilize inference**

Some of these are hard to avoid…e.g. TMLE with Super Learner requires choices about the candidate learners, their tuning parameters, the cross-validation scheme and loss function

Part 3 conclusions

- Current covariate adjustment methods for ITCs make strong parametric assumptions and may be subject to bias
- ML-NMR is clearly relevant for EU HTA: with limited subject-level data, it is the only method that can provide treatment effect estimates in any target population
- Doubly robust data-adaptive covariate adjustment approaches show promise…but it can be challenging to integrate these within larger networks of treatments and studies
- While data-adaptive estimation can automate certain researcher degrees of freedom, it also introduces others, and carries its own set of challenges with respect to transparency

4. Conclusion

Some suggestions for future steps

Concluding remarks

The EU HTA Regulation will result in:

- Greater analytical complexity with respect to the number and type of ITCs being conducted
- A widespread need for covariate-adjusted ITCs, to address highly variable PICOs across specific target (sub) populations

There is an urgent need:

- For the development and promotion of more bias-robust covariate-adjusted ITC methods
- To upskill and train statisticians in advanced ITC methods, across industry and HTA bodies, to keep pace with methodological innovation
- To strengthen and update best-practice recommendations and reporting guidelines more regularly
- To expand capacity and statistical resources dedicated to ITCs within industry
- To enhance capabilities in ITC methodology within the HTA staff and committee members assessing the evidence

Concluding remarks

Careful and prospective ITC planning/pre-specification at the trial design stage:

- As part of a HTA-specific statistical analysis plan for registrational trials
- To think clearly about researcher degrees of freedom in the methodology implementation
- To improve transparency
- To elicit and measure important prognostic factors and effect modifiers

- **Requires workable and predictable PICOs….**
- **…derived transparently using evidence-based methods, and engaging with the health technology developer throughout the process!**

References

- Ades, A.E., Welton, N.J., Dias, S., Phillippo, D.M. and Caldwell, D.M., 2024. Twenty years of network meta-analysis: Continuing controversies and recent developments. Research Synthesis Methods
- Cassidy, O., Harte, M., Trela-Larsen, L., Walsh, C., White, A., McCullagh, L. and Leahy, J., 2023. A comparison of relative-efficacy estimate (s) derived from both Matching-Adjusted Indirect Comparisons and Standard Anchored Indirect Treatment Comparisons. Value in Health.
- EUnetHTA 21 Practical Guideline D4.2 Scoping Process, Version 1.0, 12/09/2022
- Jackson, D., Zhang, L., Hettle, R. and Cooper, M., 2024. 'Twenty years of network meta-analysis: Continuing controversies and recent developments': A health technology assessment perspective. Research Synthesis Methods.
- Macabeo, B., Rotrou, T., Millier, A., François, C. and Laramée, P., 2024. The Acceptance of Indirect Treatment Comparison Methods in Oncology by Health Technology Assessment Agencies in England, France, Germany, Italy, and Spain. PharmacoEconomics-Open, 8(1), pp.5-18.
- 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
- Serret-Larmande, A., Zenati, B., Dechartres, A., Lambert, J. and Hajage, D., 2023. A methodological review of population-adjusted indirect comparisons reveals inconsistent reporting and suggests publication bias. Journal of Clinical Epidemiology.
- Truong, B., Tran, L.A.T., Le, T.A., Pham, T.T. and Vo, T.T., 2023. Population adjusted-indirect comparisons in health technology assessment: A methodological systematic review. Research Synthesis Methods, 14(5), pp.660-670.
- van der Laan, M., 2024. Targeted Learning for Causal Inference Using Real-World Data. ISPOR 2024.
- van Engen, A., Krüger, R., Parnaby, A., Rotaru, M., Ryan, J., Samaha, D. and Tzelis, D., 2024. The Impact of Additive PICOs in a European Joint Clinical Health Technology Assessment. Value in Health.

Audience Q&A