EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

Bias in indirect treatment comparisons and evolving methodology: implications for health technology assessment and beyond

Brought to you by the HTA ESIG

Session Chair: Katrin Kupas, BMS

"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."

Visit Us

EFSPI EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

Opening Remarks

"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."

Session Overview

EFSPI EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

Session Open Katrin Kupas, BMS	**Meeting Evidence Requirements in the EU HTA Landscape: PICOs and ITCs** Dave Gelb, MSD	**Methodologies to adjust for measured confounding in ITC: an overview of population adjustment approaches** David Philippo, University of Bristol
Methodologies to adjust for unmeasured confounding in ITC Kate Ren, University of Sheffield	**Case Study** Nicolas Scheuer, Roche	**Panel Discussion & Close** Katrin Kupas, BMS

"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."

PICOs & ITCs

Meeting Evidence Requirements in the EU HTA Landscape

18th June 2024

Dave Gelb, Lara Wolfson, Audrone Aksomaityte

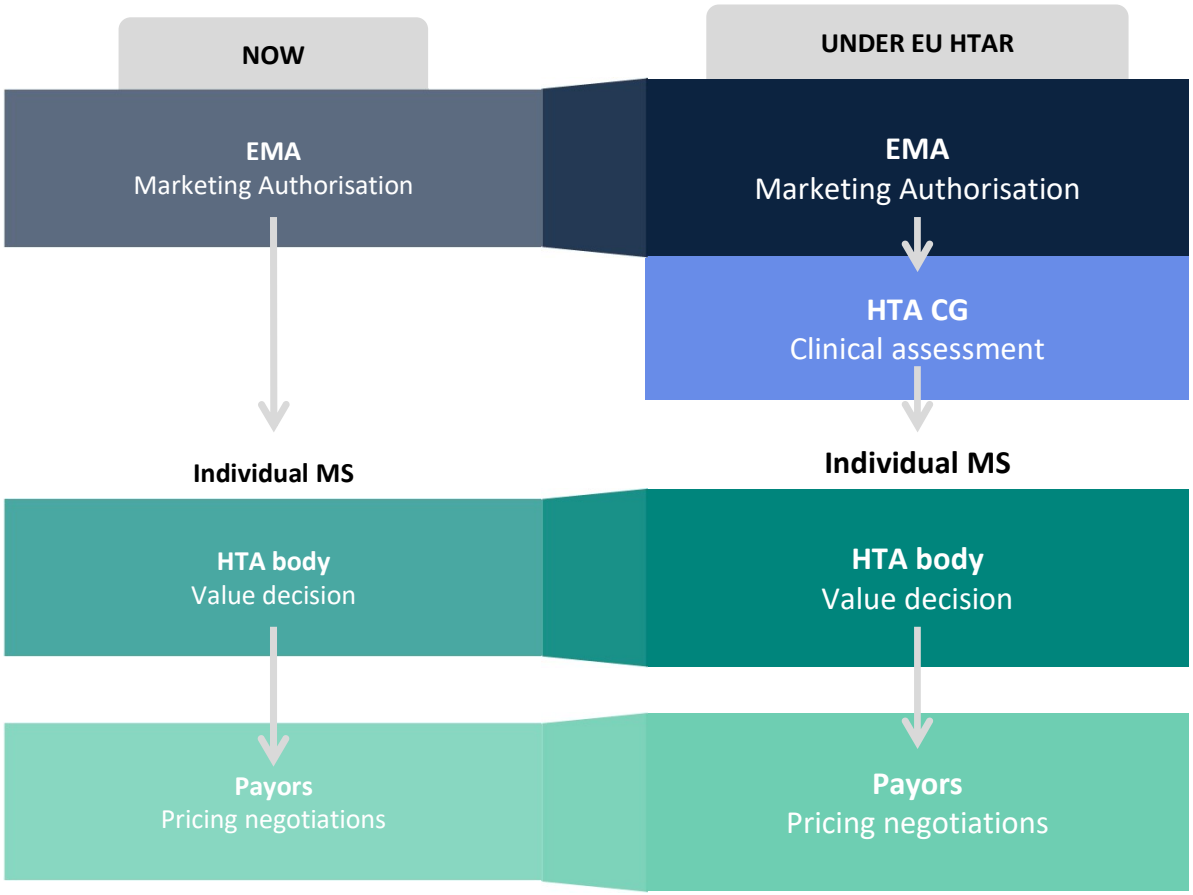
HTA Statistics, BARDS HTA Statistics, MSD, Zurich Switzerland

Disclaimer: All views expressed in this talk are my own and do not necessarily represent the views of my employer

What does implementation of the EU HTAR trigger?

The mandatory requirement of centralised clinical assessment for patient access in Europe

- An HTA assessor and co-assessor are appointed;
- They determine the scope of the assessment (PICOs, Population, Intervention, Comparator, Outcomes)
- The manufacturer is informed around ~3 months after EMA submission of the selected PICOS
- The HTA dossier is submitted no later than 45 days before CHMP opinion
- A *"***J**oint **C**linical **A**ssessment"(JCA) report is available within 30 days of market authorization



CG, Coordination Group; EMA, European Medicines Agency; EU, European Union; HTA, Health Technology Assessment; HTAR, Health Technology Assessment Regulation; MS, Member State.

Understand your PICOs

The assessment scope for EU HTA will be PICO - based :

- **P**atient population
- **I**ntervention
- **C**omparator(s)
- **O**utcomes

PICO selection is policy-driven, not evidence-driven.

Each Member State declares their target PICO(s). These PICOs are then consolidated, approximately 90-140 days after regulatory submission

CONSEQUENCES

- Expect more comparisons than in PhIII trials to be requested, including a substantial volume of indirect treatment comparisons
- Data will become publicly available shortly after regulatory approval
- May constrain national-level economic models
- May impact treatment guidelines at national level

CHALLENGES

- HTD is not involved in PICO determination process (sits at EU level), consultation meeting possible
- PICOs must be estimated internally by HTD for JCA dossier planning
- High number of PICOs may be requested for JCA
- Adaptations of PICOs may be requested in response to regulatory discussion

EU, European Union; HTD, Health Technology Developer; JCA, Joint Clinical Assessment; MS, Member State; PICO, Population, Intervention, Comparator(s), Outcomes.

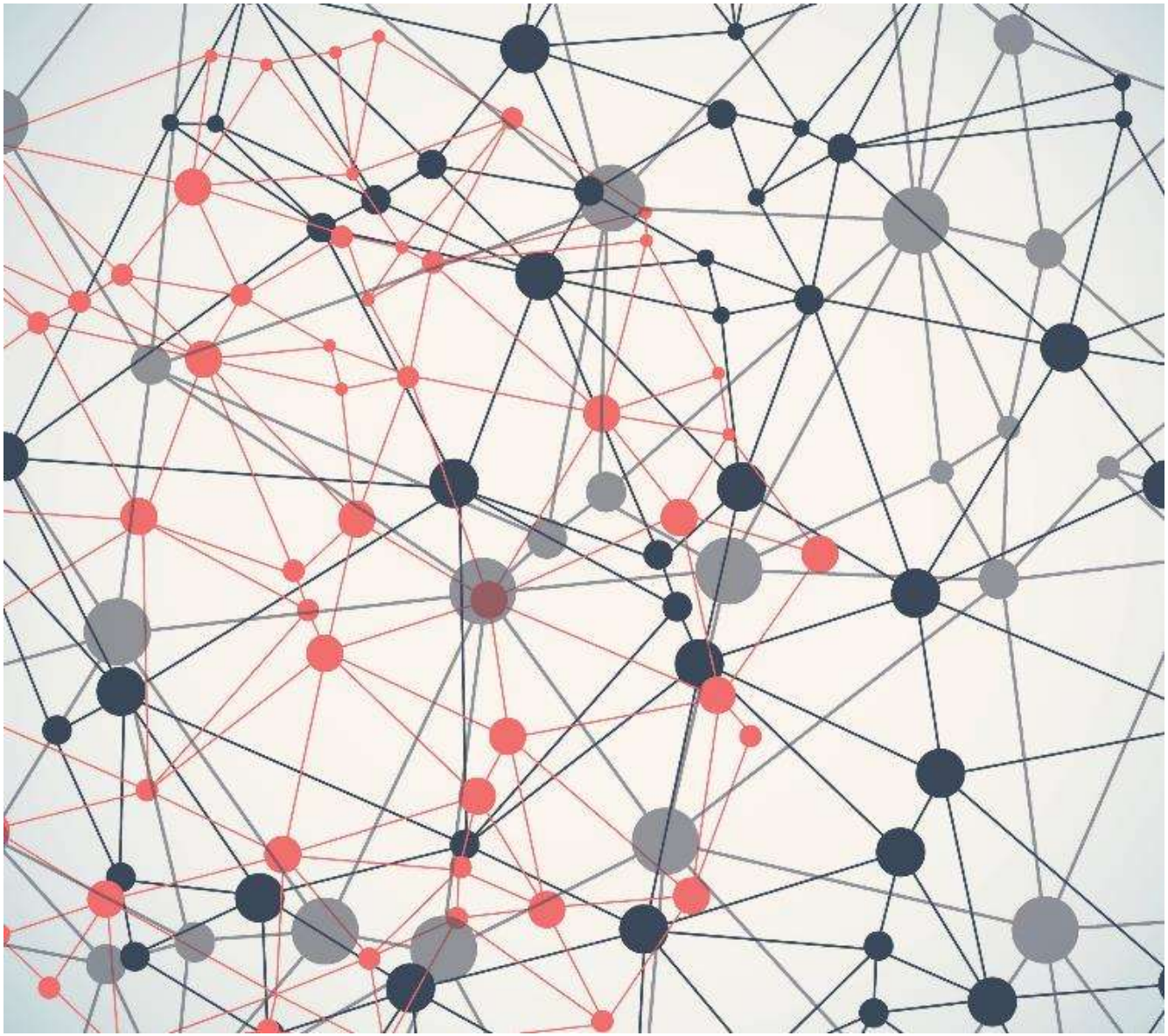
How many PICOs?

EFPIA-Evidera Simulation of EU HTA JCA Process for 3 Oncology Products*:

Category	Product X	Product Y	Product Z
Populations	2	10	10
Comparators	15	8	23
Outcome Categories	5	7	5
Consolidated PICOs	7-16	6-22	23-57

*"Meta-analysis and ITCs will be critical to meet the evidence development requirements of likely multiple PICOs outlined in a JCA scope"**

* Source: https://www.efpia.eu/media/qrjah2ij/efpia-evidera-research-on-eunethta21-methods.pdf

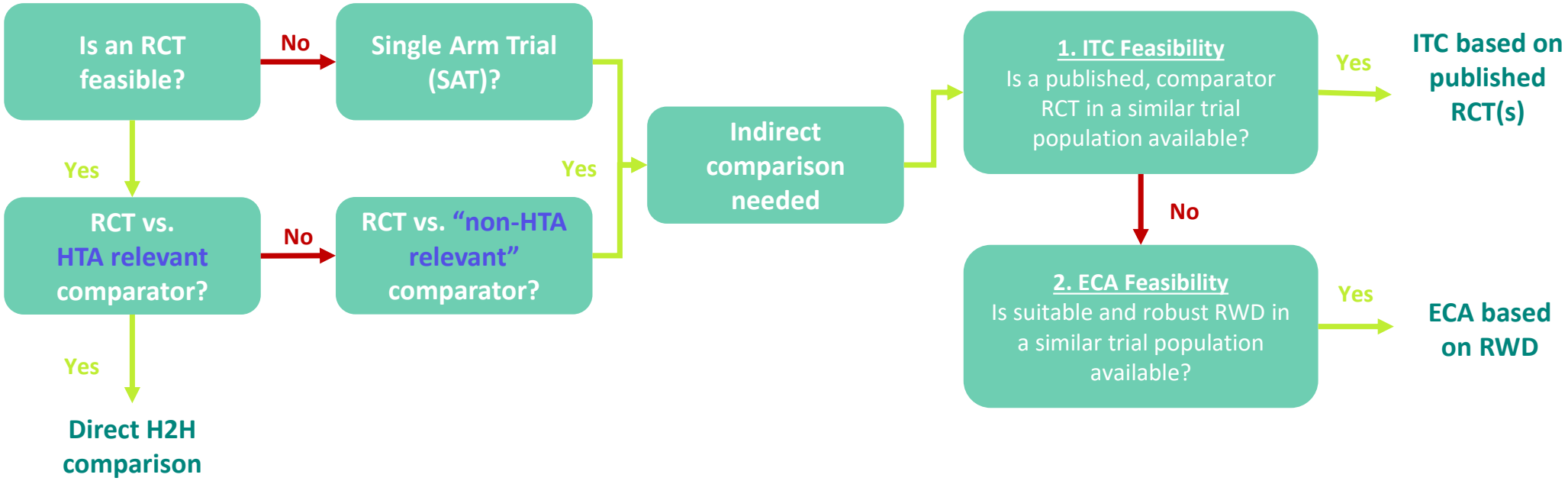


What are Indirect Treatment Comparisons?

Why Use Indirect Treatment Comparisons?

- HTA seeks to understand **comparative clinical effectiveness**
- RCTs are "gold standard" for direct comparisons but don't always include all relevant comparators
- If no indirect evidence available, then RWD can potentially be used to address the evidence gap

Comparative Evidence Options for HTA submissions



Abbreviations: RCT: Randomized Controlled Trial, ITC: Indirect treatment comparison, ECA: External comparator arm; RWD: Real-World Data; SAT: Single-Arm Trial; H2H: Head-2-Head
Source: Adapted from IQVIA Report to MSD on ITC Usage in European Health Technology Assessments, March 2024

What are ITCs?

- A “library” of methods; ITC, NMA, STC, MTC, MAIC, Bucher, ML-NMR, IPTW, Lumley, Bayesian NMA, …
- Whenever you don’t have direct treatment evidence against the comparator of interest

Indirect treatment comparison/mixed treatment comparison	The estimation of the relative effectiveness of two or more treatments in the absence of any head-to-head trials
Mixed treatment comparison/network meta-analysis	The simultaneous estimation of the relative effectiveness of three or more treatments using a combination of direct and indirect evidence and a common comparator
Bucher's adjusted indirect comparison	Adjusted indirect method of treatment comparison that can estimate relative treatment effects for star pattern networks
Population adjusted ITC	Indirect treatment comparison in which IPD in one or more trials are used to adjust for between-trial differences in the distribution of variables that influence outcome
Matching-adjusted indirect comparison	A form of propensity score weighting, applicable where IPD are available in one population and aggregate data in another
Simulated treatment comparison	A form of outcome regression, applicable where IPD are available in one population and aggregate data in another
Naive indirect comparison	Comparison of competing clinical interventions from data of individual arms of different studies, based on the assumption that the treatment groups are clinically homogeneous in composition

IPD: Individual patient data; ISPOR: International Society for Pharmaco-economics and Outcomes Research; ITC: Indirect treatment comparison.

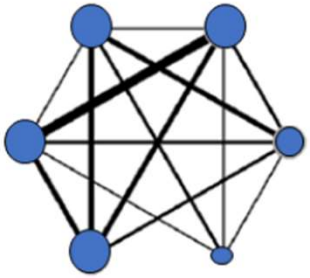
Source: European Network for Health Technology Assessment. Comparators & Comparisons: Direct and indirect comparisons (2013) and Conducting Indirect-Treatment-Comparison and Network-Meta-Analysis Studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices – Part 2.

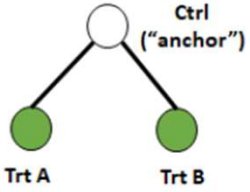
Source: Skali & Spoors 2017 https://becarispublishing.com/doi/10.2217/cer-2017-0092

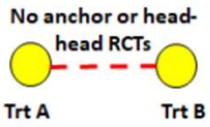
ITCs are all about the choices…and what's available

Types of available data
(aggregate, patient level, or both?)

Evidence structure
(anchored? disconnected?)







Number of therapies to be compared
(e.g., two? many?)

Population heterogeneity
(many differences between study populations?)

Source: https://www.linkedin.com/pulse/choosing-between-methods-indirect-treatment-brian-hutton/

How do HTA Bodies View ITCs?

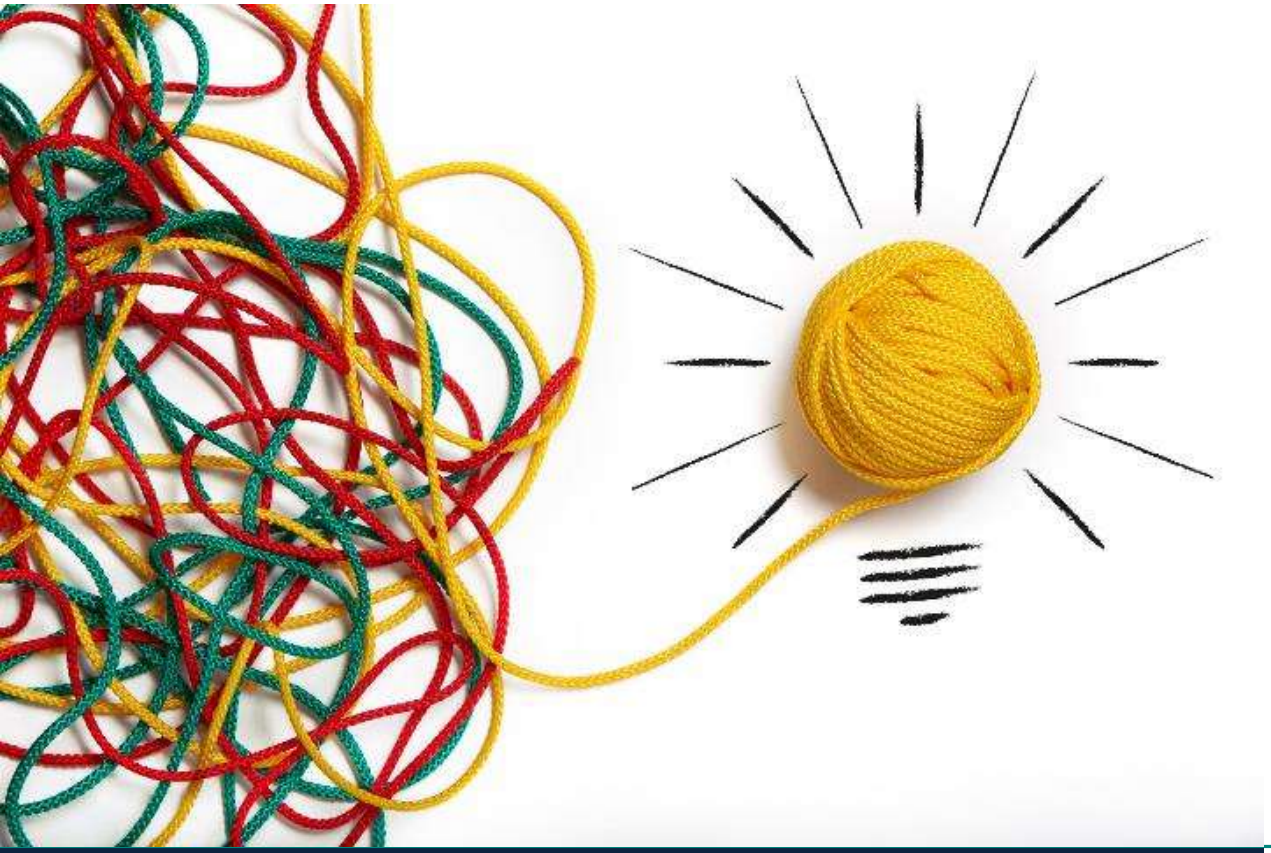
Acceptability of methods varies across HTA agencies

Method	EUHTA	NICE	IQWIG	PBAC	CADTH	HAS	ICER
Bucher ITC	Yes	Yes	Yes	Yes	Yes	No	Unknown[3]
MAIC/STC	Yes	Yes	Potentially	Yes	Yes	Yes	Unknown[3]
Bucher NMA	Yes	Potentially[1]	No/Potentially[2]	No	Yes	No	Unknown[3]
Frequentist NMA (Lumley)	Yes	Potentially[1]	No/Potentially[2]	No	Yes	Yes	Unknown[3]
Bayesian NMA	Yes	Yes	No/Potentially[2]	No	Yes	Yes	Yes

1: NICE has clear preference for Bayesian NMAs, but could consider frequentist approaches if assumptions are satisfied
2: IQWIG does not endorse NMAs but could accept it depending on the research question
3: No statement has been made about those methods

NICE: National Institute for Health and Care Excellence (HTA agency in United Kingdom); IQWiG: German Institute for Quality and Efficiency in Health Care; PBAC is Pharmaceutical Benefits Advisory Committee (HTA advisory in Australia); CADTH: Canada's Drug and Health Technology Agency; HAS is Haute Autorité de Santé (HTA advisory in France); ICER: Institute for Clinical and Economic Review (independent health technology value assessment in the United States)
MAIC: matching-adjusted indirect comparison; STC is simulated treatment comparison; IPD is individual patient-level data; AgD is aggregate data; NMA: Network meta-analysis

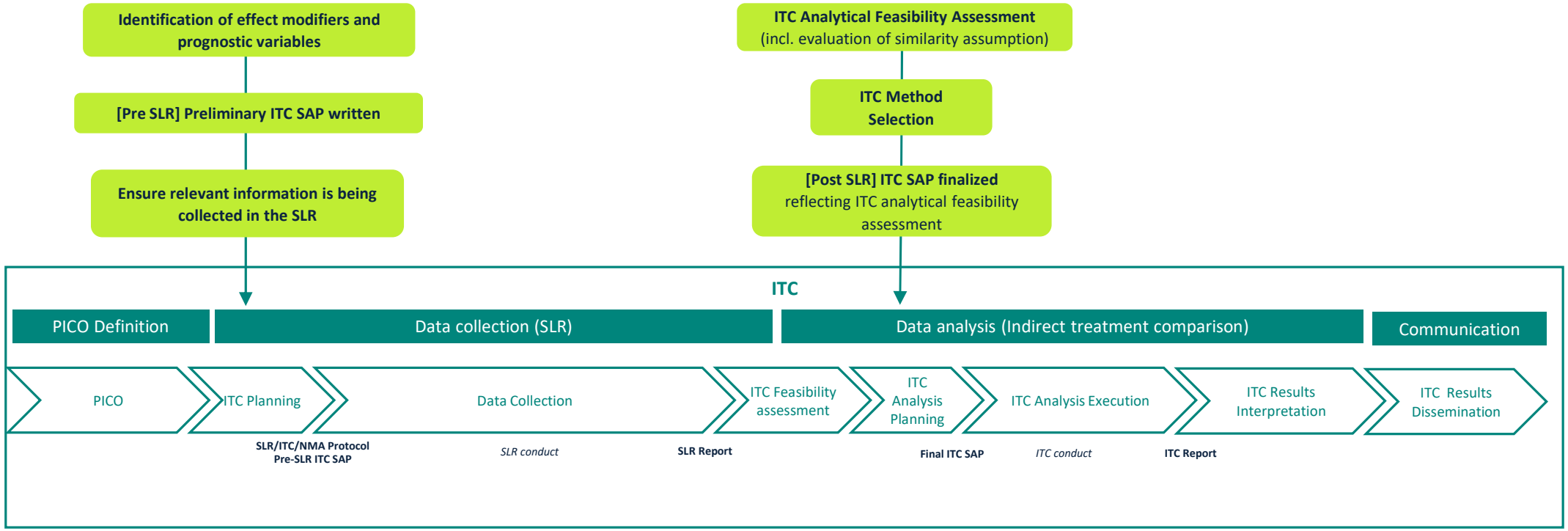
Sources:
Internal Review MSD November 2023
Member State Coordination Group on Health Technology Assessment, Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons



Are there some recommended best practices?

Key Process Steps for Conducting an ITC:
It's all about good planning for the SLR – and having an SAP

This slide outlines key process steps that should be performed for robust ITC evidence-generation to meet external requirements for impactful HTA.



Don't forget that all protocols, SAPs, and possibly programming code, are part of the EU HTA JCA Submission

BEST PRACTICE:
Identification and Pre-Specification of Effect Modifiers and Prognostic Variables

Definitions

Prognostic variables: characteristics that affect the outcome of interest irrespective of which treatment is received

Effect modifiers: characteristics that alter the relative effectiveness of an outcome between two treatments. Of note, effect modifiers may be specific to a treatment, effect measure, or trial population.

Sources and Reporting

Consider multiple sources:

1. Expert clinical opinion (external and internal) [*]
2. Results of other studies on the therapeutic indication [*] and subgroup analyses.
3. Literature search for subgroups with different treatment effects

Comprehensive and transparently reporting is key! [*]

Potential Effect Modifiers and Prognostic Variables

The following aspects should be evaluated to identify possible effect modifiers:

Study and patient characteristics: e.g., age, sex, disease severity, region, etc. [*]

Characteristics of the intervention and the comparator: e.g., dosage, application, and concomitant treatments. [*]

Best practice is *a priori* identification of effect modifiers

References:
*HTA CG; Member State Coordination Group on Health Technology Assessment. Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons
Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta-analyses. Cochrane handbook for systematic reviews of interventions, 241-284.

MSD

Key Takeaways: Patients, Payers, and Providers want to know:
In a dynamic treatment landscape, which treatment is the best choice?

The use of indirect treatment comparisons will be critical in HTA value assessment!

Pre-planning is important, and **statistical engagement** in the appropriate planning, execution, review and interpretation of ITCs is key

Different HTA bodies **have different requirements** and may prefer different methods

→ Anticipate doing the same analysis in different ways for different stakeholders

Clear and transparent disclosure of assumptions, data sources, decisions, and sensitivity analyses will be key to building trust and transparency

MSD
INVENTING FOR LIFE

Thank you

Acknowledgements: Some of the material in this presentation was supported by contributions from MSD staff members Mikkel Oestergaard, Lauren Abderhalden, Abdel Hmissi, Robson Machado, Dominic Muston, Justin Chumbley, Lidia Mukina, Rachid Massaad, Jo Gregory, Gregory Chen, Shahrul Mt-Isa, Céline Le Bailly De Tilleghem
Views in this presentation are those of the presenter and not of the company as a whole.

Methodologies to adjust for measured confounding in ITC: an overview of population adjustment approaches

David Phillippo

Bristol Medical School (Population Health Sciences)

University of Bristol, UK

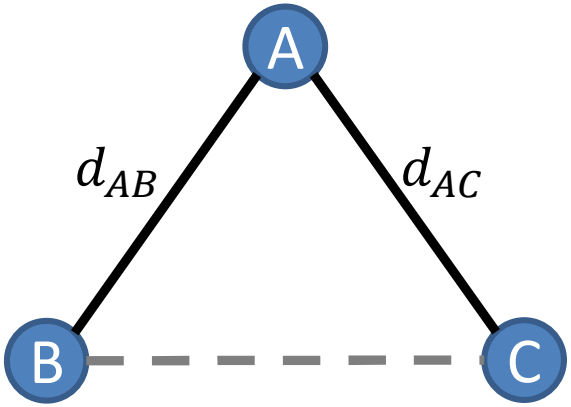
University of BRISTOL

Background

We wish to compare multiple treatments, but not all are studied in the same trial

Standard methods using aggregate data (AgD):

- Indirect comparison: $d_{BC} = d_{AC} - d_{AB}$
- Network meta-analysis (NMA)



- Assume constancy of relative effects:

$$d_{AB(AB)} = d_{AB(AC)}$$

- Biased if there are differences in effect modifiers between studies

Background – population adjustment

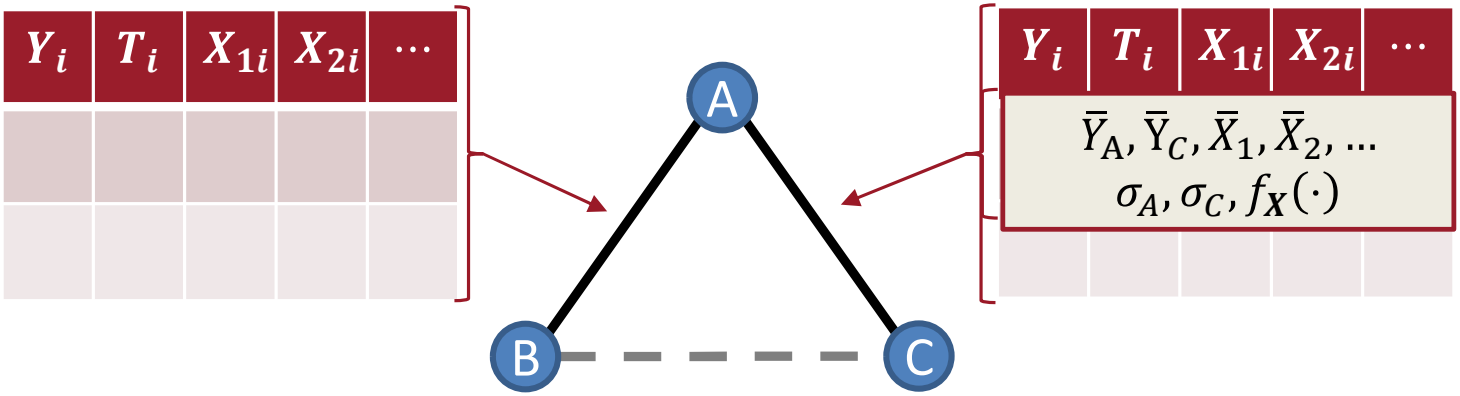
Population adjustment methods make use of available individual patient data (IPD) to adjust for effect modifiers

Ideal scenario: full IPD

- "Gold standard" is IPD meta-regression

Common scenario: limited IPD

- Several recent methods make use of mixed data

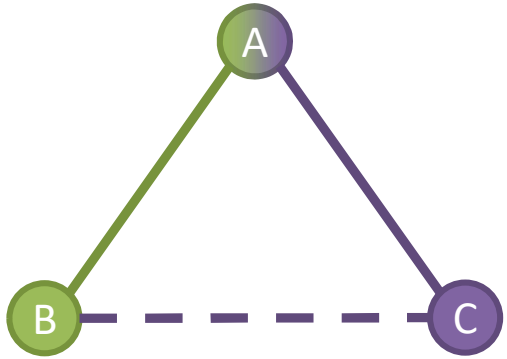


University of BRISTOL

Anchored vs. Unanchored Comparisons

Anchored population-adjusted indirect comparisons

- Common comparator, respect randomisation
- Assume conditional constancy of relative effects
- Predict $d_{AB(AC)}$ from the AB trial
- All effect modifiers known and adjusted for



Unanchored population-adjusted indirect comparisons

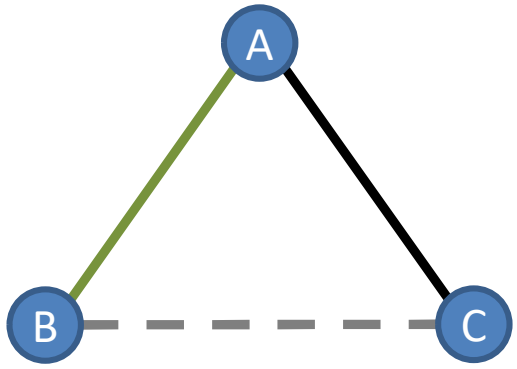
- No common comparator, no randomisation
- Assume conditional constancy of absolute effects
- Predict $Y_{B(C)}$ from the B trial
- All effect modifiers and prognostic variables known and adjusted for

Population adjustment – MAIC and STC

Matching-Adjusted Indirect Comparison	**Simulated Treatment Comparison**
• Population reweighting method	• Outcome regression method
• Weight AB individuals to balance covariate distribution with AC trial	• Fit regression model in AB trial
• Estimate outcomes on A and B in AC trial using weights	• Estimate outcomes on A and B in AC trial using regression model

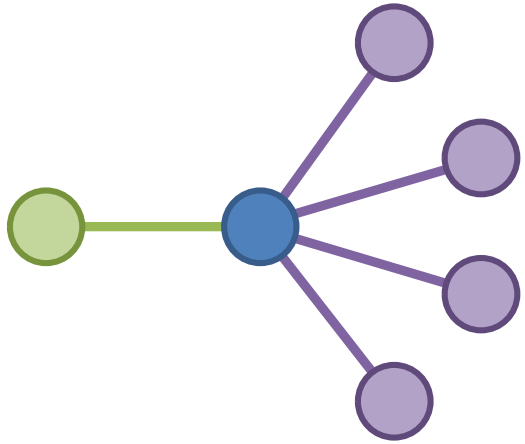
Limitations

- Limited to pairwise indirect comparisons
- Comparisons stuck in aggregate (AC) population
- STC can incur aggregation bias with non-linear models, non-collapsibility bias



Multiple comparators – problematic for MAIC and STC

- Larger networks are already commonplace in HTA
  - 2019 review of NICE TAs with population-adjustment found 56% involved larger networks
- Likely to increase with JCA
  - Required to consider more comparators
- MAIC and STC cannot handle larger networks
  - Multiple analyses are incoherent, re-use the data
  - Each analysis valid for a different target population



Multilevel Network Meta-Regression (ML-NMR)

Phillippo et al. (2020)

- Applicable in networks of all sizes
- Avoids aggregation bias
- Correctly handles non-collapsible effect measures
- Produces estimates in any target population for decision making
- Extends the standard network meta-analysis (NMA) framework, reducing to:
  - IPD network meta-regression with full IPD
  - Standard NMA with no adjustment
- Allows assumptions to be tested/relaxed in larger networks (Phillippo et al. 2023)
- Implemented in R package *multinma*

ML-NMR

1. Define an individual-level regression model
   - IPD network meta-regression (gold-standard approach)
2. Average (integrate) this over the aggregate population(s) to form the aggregate-level model

A generalised form of this approach can be applied to survival outcomes (Phillippo et al. 2024)

Individual:

$$y_{ijk} \sim \pi_{\text{Ind}}(\theta_{ijk})$$
$$g(\theta_{ijk}) = \eta_{jk}(\boldsymbol{x}_{ijk}) = \mu_j + \boldsymbol{x}_{ijk}^{\mathsf{T}}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,k}) + \gamma_k$$

Aggregate:

$$y_{\bullet jk} \sim \pi_{\text{Agg}}(\theta_{\bullet jk})$$
$$\theta_{\bullet jk} = \int_{\mathfrak{x}} g^{-1}\left(\eta_{jk}(\boldsymbol{x})\right) f_{jk}(\boldsymbol{x})\, d\boldsymbol{x}$$

Use numerical integration

Predicting quantities of interest for a target population

The target population could be represented by

- A randomised trial
- An observational study
- A registry dataset
- …

With IPD covariate information

1. Make predictions for each individual
2. Summarise these for the population

With summary statistics

1. Generate integration points from joint covariate distribution
2. Integrate over the target population

Predicting quantities of interest for a target population

Population-average **conditional** treatment effects, simplify to "plugging-in" mean covariate values:

$$d_{ab(P)} = \bar{\boldsymbol{x}}_{(P)}^{\mathsf{T}}(\boldsymbol{\beta}_{2,b} - \boldsymbol{\beta}_{2,a}) + \gamma_b - \gamma_a$$

Absolute predictions (e.g. average event probabilities):

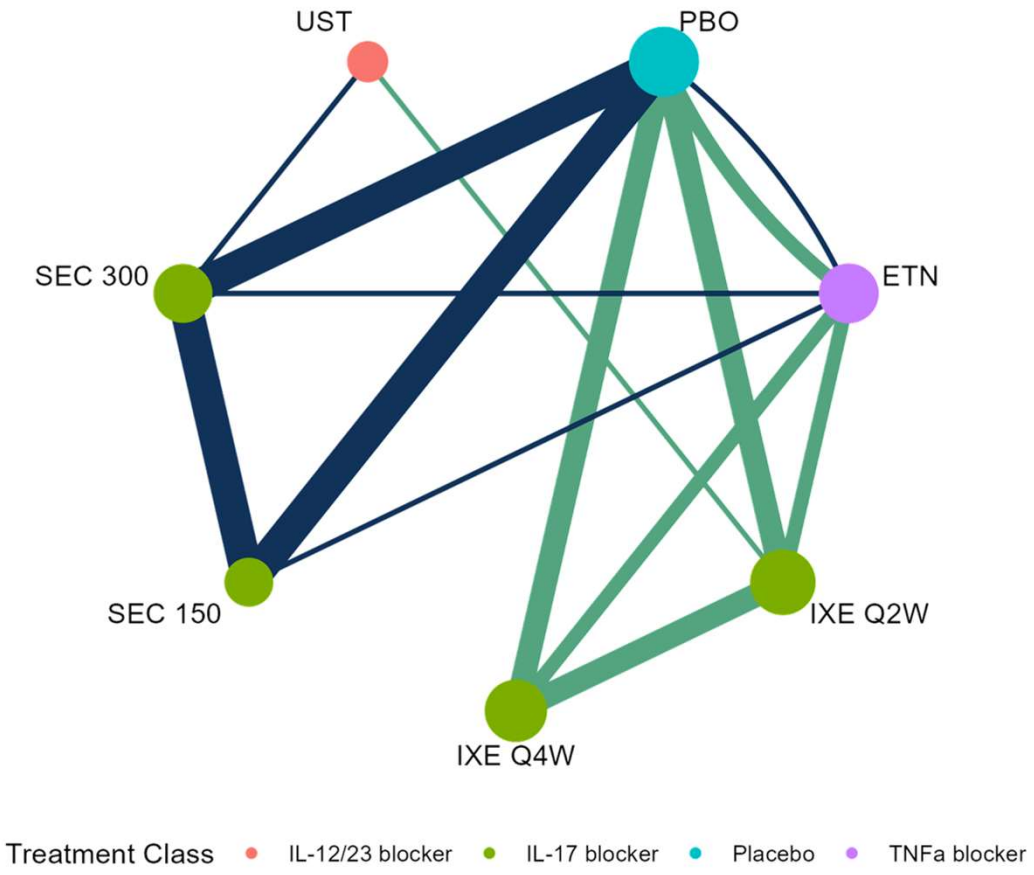
$$\bar{p}_{k(P)} = \int_{\mathfrak{x}} g^{-1}\left(\mu_{(P)} + \boldsymbol{x}^{\mathsf{T}}\left(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,k}\right) + \gamma_k\right) f_{(P)}(\boldsymbol{x})\, d\boldsymbol{x}$$

Population-average **marginal** treatment effects:

$$\Delta^{\mathrm{RD}}_{ab(P)} = \bar{p}_{b(P)} - \bar{p}_{a(P)} \qquad \Delta^{\mathrm{LOR}}_{ab(P)} = \mathrm{logit}\left(\bar{p}_{b(P)}\right) - \mathrm{logit}\left(\bar{p}_{a(P)}\right)$$

Example: Plaque Psoriasis

Phillippo et al. (2020)

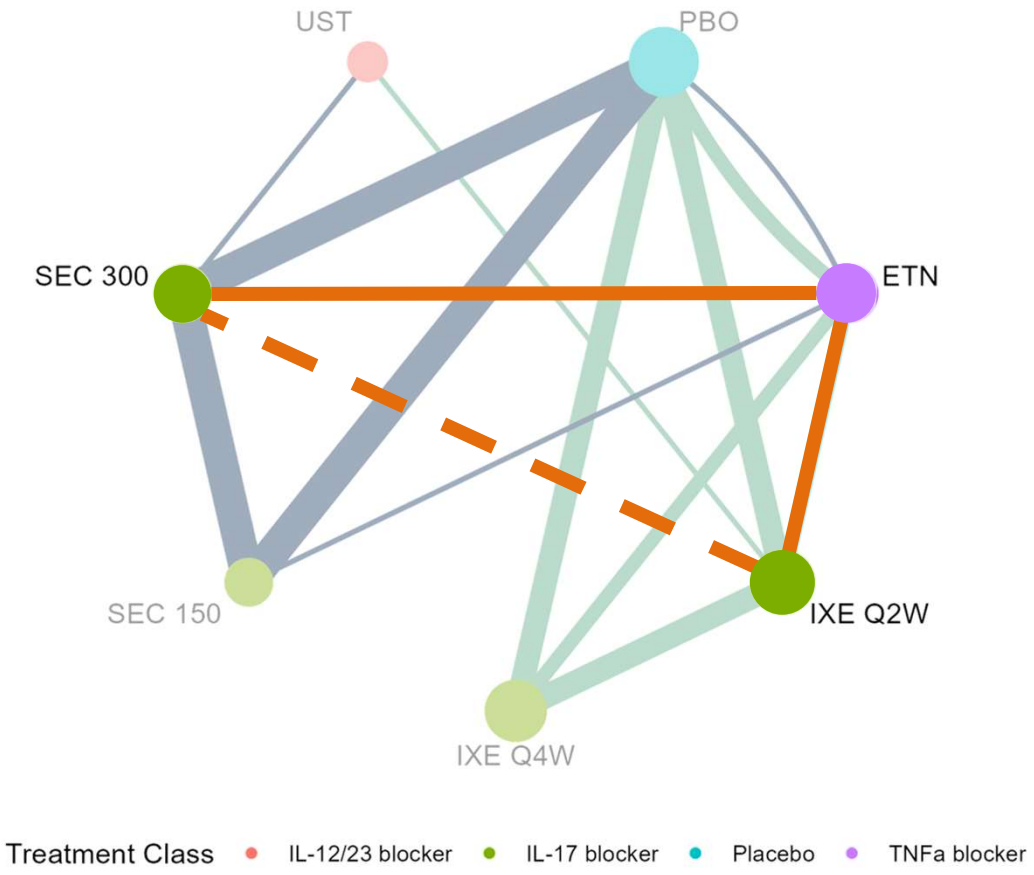


- Seven treatments for plaque psoriasis
- Four **IPD** studies
- Five **AgD** studies
- Outcomes are binary response on PASI scale (75%, 90%, 100%)
- Five potential effect modifiers to adjust for
  - Previous systemic treatment
  - Duration of psoriasis
  - Body surface area covered
  - Weight
  - Psoriatic arthritis
- External target population - PROSPECT registry

University of BRISTOL

Example: Plaque Psoriasis

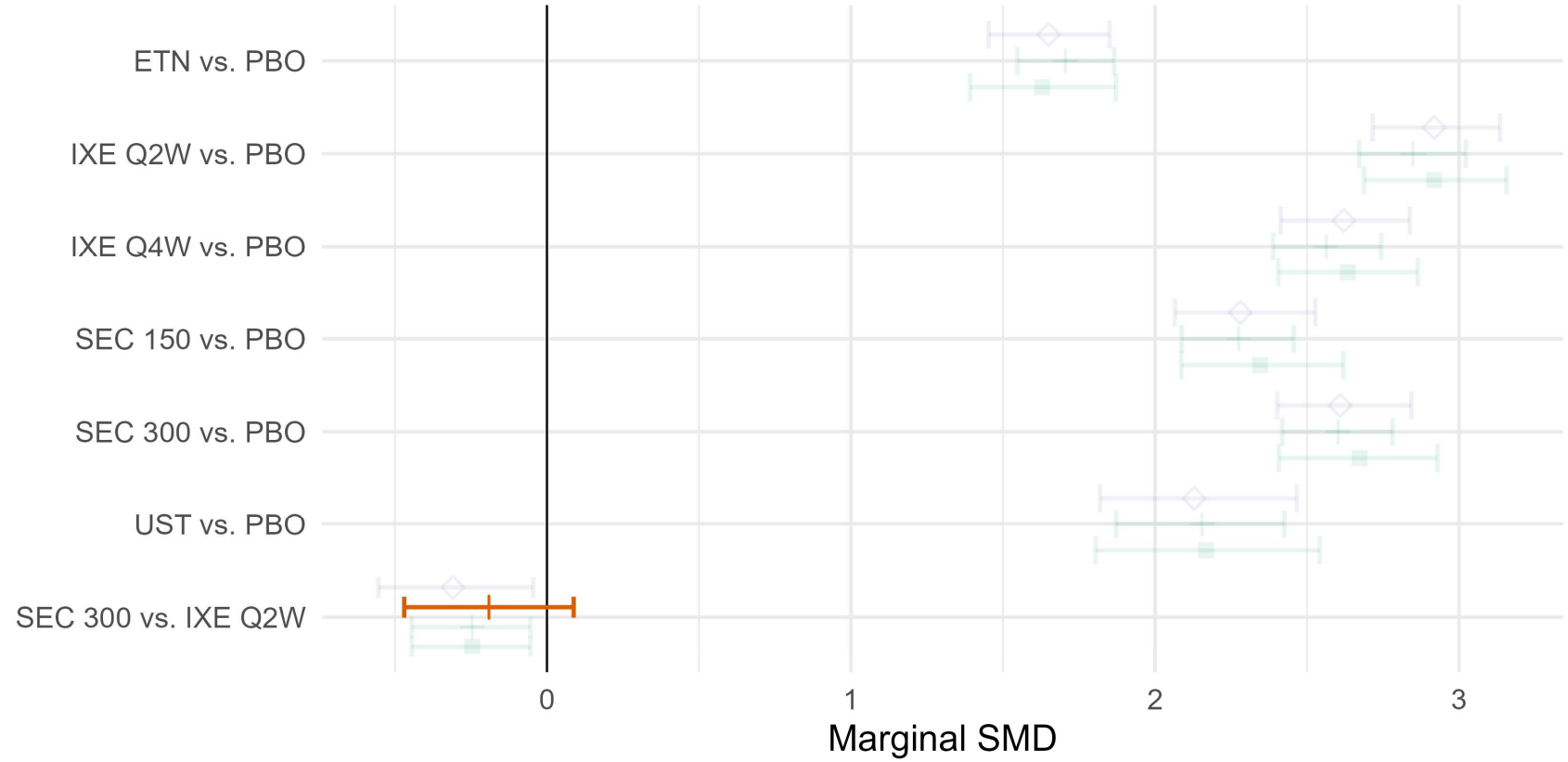
Phillippo et al. (2020)

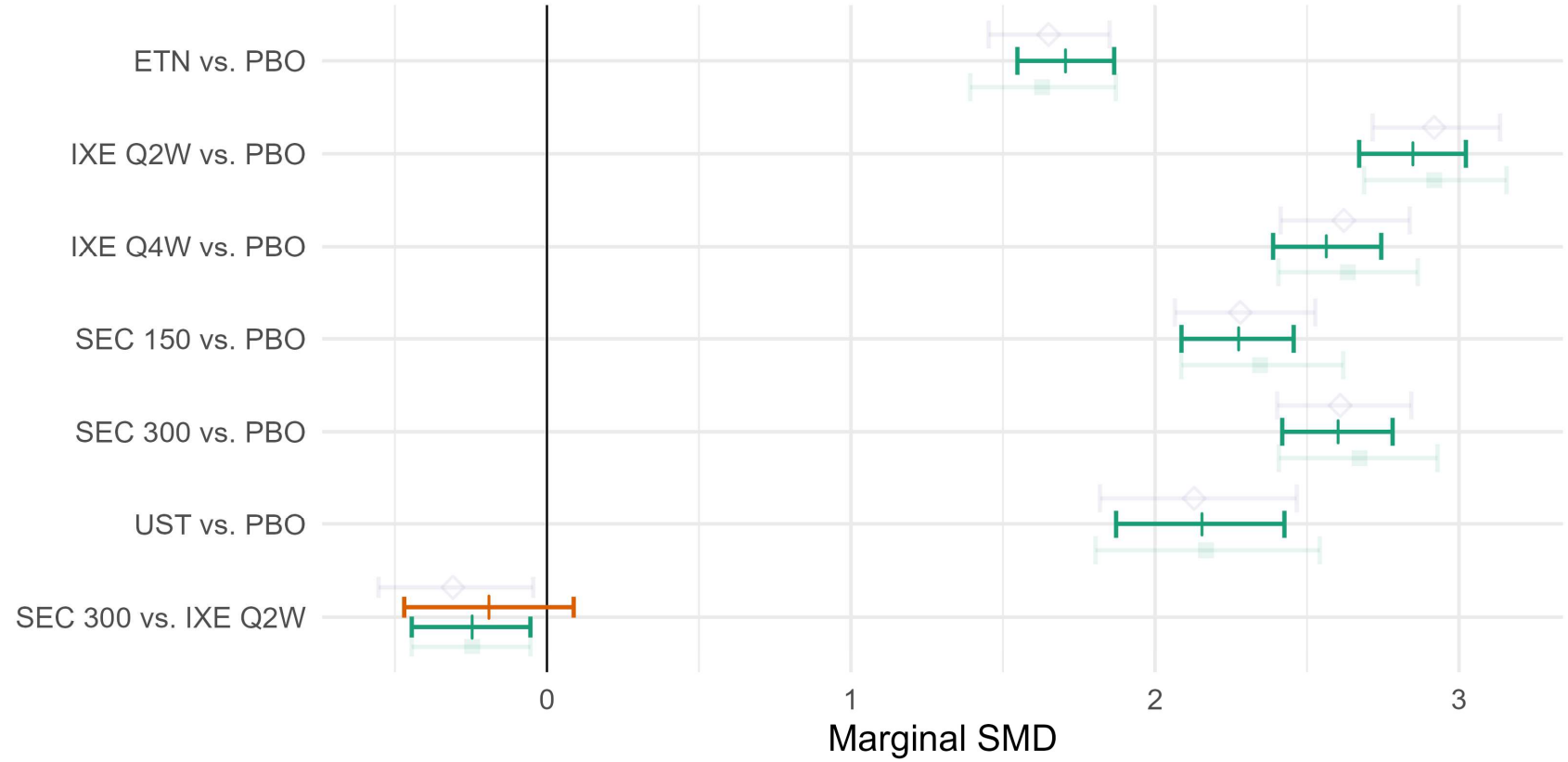


Previous MAIC compared IXE Q2W and SEC 300 via ETN

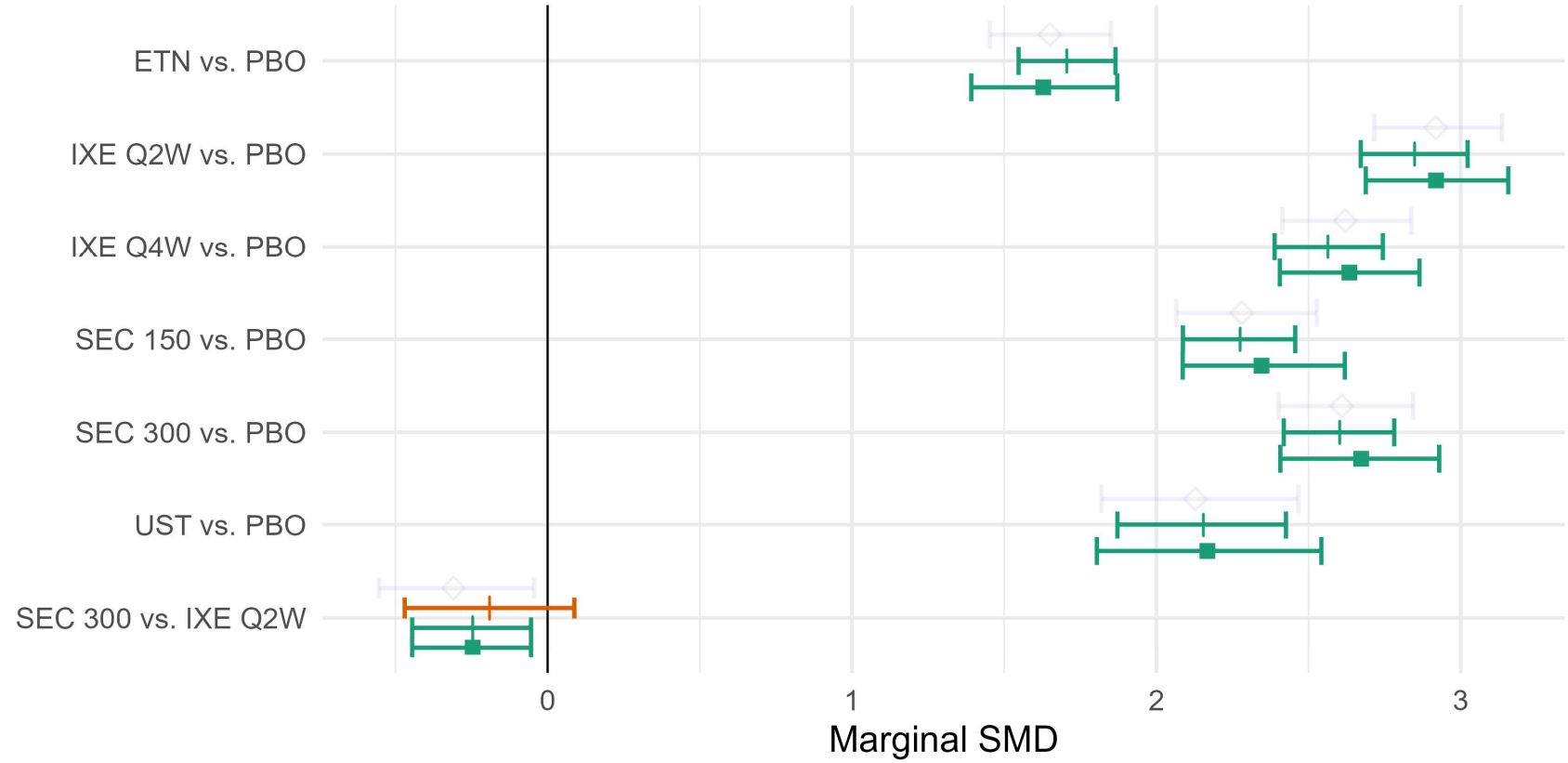
- Could have used common comparator PBO instead
- Could not use information from one IPD study, four AgD studies
- Estimates only in single aggregate population (FIXTURE)
- Unable to obtain a coherent set of effect estimates for all treatments

University of BRISTOL

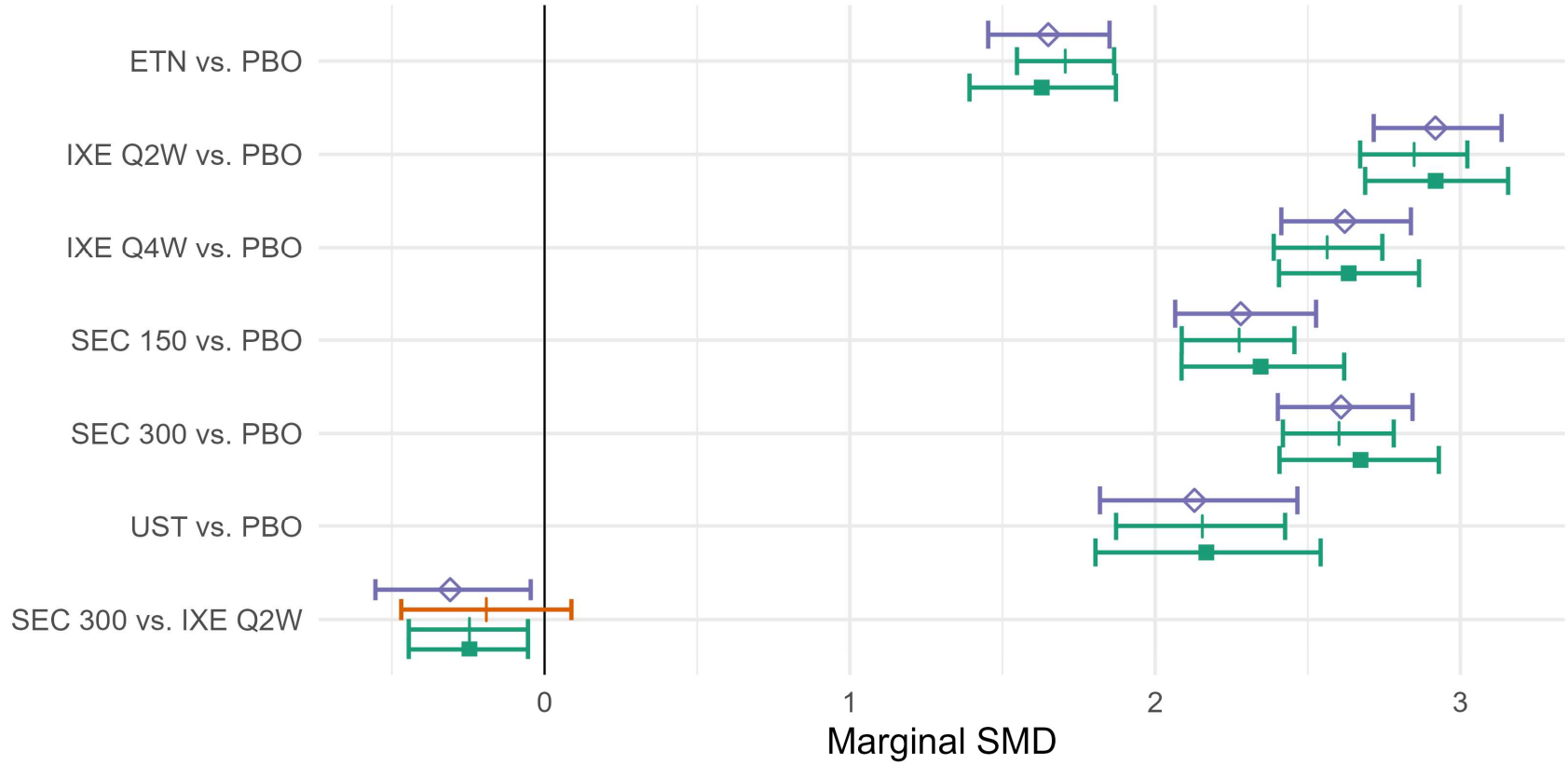

ETN vs. PBO
IXE Q2W vs. PBO
IXE Q4W vs. PBO
SEC 150 vs. PBO
SEC 300 vs. PBO
UST vs. PBO
SEC 300 vs. IXE Q2W
0
1
2
3
Marginal SMD
Method
ML-NMR
MAIC
RE NMA
Study population
RE NMA
FIXTURE
PROSPECT



- Produce a full set of coherent estimates
- Reduced uncertainty compared to MAIC



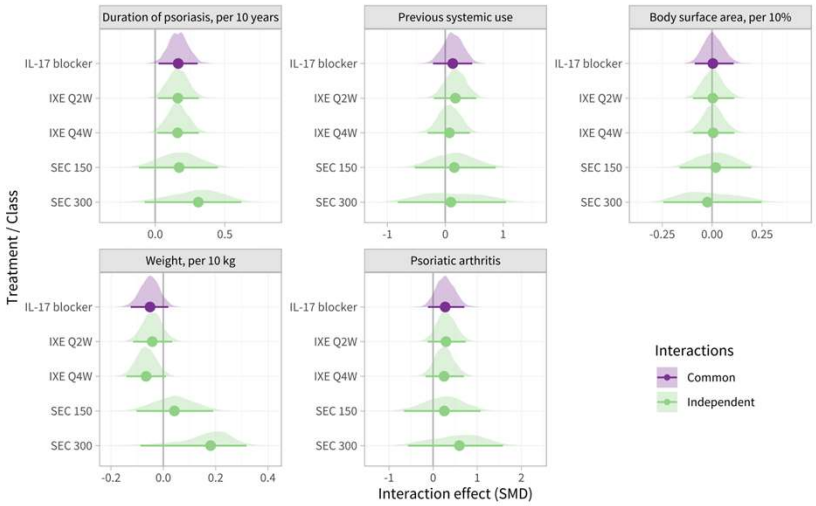
- Produce a full set of coherent estimates - in any target population
- Reduced uncertainty compared to MAIC



- Produce a full set of coherent estimates - in any target population
- Reduced uncertainty compared to MAIC
- Reduced uncertainty compared to RE NMA

Assessing assumptions with ML-NMR

- Violation of conditional constancy (e.g. unobserved effect modifiers) may be detected using standard NMA methods
  - Random effects models – residual heterogeneity
  - Inconsistency models – residual inconsistency
  - None detected in plaque psoriasis example
- Shared effect modifier assumption may be relaxed, one covariate at a time in smaller networks

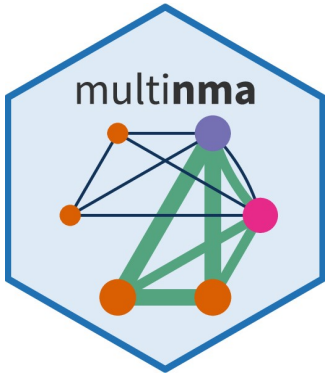


Phillippo et al. (2023)

University of BRISTOL

Summary

- ML-NMR is a flexible and general method for synthesising evidence from mixtures of individual and aggregate level data
- Several advantages over previous population-adjustment methods
  - Coherently analyse networks of any size
  - Produce estimates in a relevant decision target population
  - Assess key assumptions in larger networks
- Implemented in multinma R package
  - Website: dmphillippo.github.io/multinma
  - Documentation, example analyses



University of BRISTOL

Funding and References

This work was undertaken with the support of the MRC grants MR/P015298/1, MR/R025223/1, and MR/W016648/1.

Phillippo, DM et al. (2020) Multilevel Network Meta-Regression for population-adjusted treatment comparisons. *Journal of the Royal Statistical Society: Series A*, 183(3):1189-1210. DOI: 10.1111/rssa.12579.

Phillippo, DM et al. (2020) Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. *Statistics in Medicine*, 39(30):4885-4911. DOI: 10.1002/sim.8759

Phillippo DM et al. (2023) Validating the assumptions of population adjustment: application of multilevel network meta-regression to a network of treatments for plaque psoriasis. *Medical Decision Making*, 43(1):53-67. DOI: 10.1177/0272989X221117162.

Phillippo DM et al. (2024) Multilevel network meta-regression for general likelihoods: synthesis of individual and aggregate data with applications to survival analysis. Preprint, *arXiv*:2401.12640 [stat.ME]

Phillippo, DM (2019) *Calibration of treatment effects in network meta-analysis using individual patient data.* PhD Thesis, University of Bristol. Available from research-information.bristol.ac.uk.

R package *multinma*, see dmphillippo.github.io/multinma for details

University of Sheffield

NIHR | National Institute for Health Research

Methodologies to adjust for unmeasured confounding in ITC

PSI conference 2024

Kate Ren, PhD
Senior Research Fellow, University of Sheffield
Director of Statistics, ConnectHEOR

NIHR Advanced Fellow
NICE Technology Appraisal Committee Member

18 June 2024

Disclaimers

- The findings and views expressed in this presentation are those of the presenter, who is responsible for its contents.
- The findings and views expressed should not be understood or quoted as being made on behalf of
  - NICE Technology Appraisal Committee,
  - National Institute for Health and Care Research (NIHR).

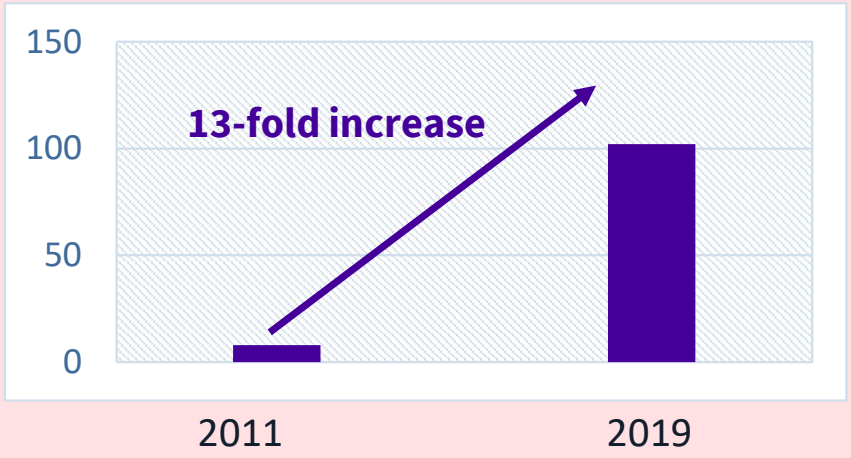
Overview

- Single-arm trials in HTA submissions
  - Confounding issue
- Quantitative bias analysis (QBA)
  - Unmeasured confounding in ITC
- Case study
  - Metastatic colorectal cancer

Single-arm trials in HTA submissions

Review of HTA submissions (2011-2019)*

433 single-arm trials



*Patel et al. (2021) doi:10.1016/j.jval.2021.01.015

Analysis methods

Population-adjusted indirect comparison (PAIC)

Confounding issue: guidance

EUnetHTA 21: Direct and Indirect Comparisons

"An assessment of whether the set of included covariates is likely sufficient to generate an unbiased comparison of outcomes; **quantification of the magnitude and direction of potential bias arising** from missing prognostic variables and effect modifiers in the analysis; If shifted hypothesis testing has been used, an assessment of whether this is sufficient to account for the likely **magnitude of residual bias** arising from missing covariates."

NICE DSU TSD 18

"Provide evidence that absolute outcomes can be predicted with sufficient accuracy in relation to the relative treatment effects, and present an estimate of **the likely range of residual systematic error**. If this evidence cannot be provided or is limited, then state that **the amount of bias** in the indirect comparison is likely to be substantial, and could even exceed the magnitude of treatment effects which are being estimated."

Confounding issue: what happens in practice?

A methodological systematic review of studies implementing PAICs*

Sensitivity analysis to assess the robustness of PAIC results	Statistics
No sensitivity analysis	**77 (47.5%)**
Adjusting for different sets of covariates	**55 (34.0%)**
Applying additional inclusion/exclusion criteria to the IPD study	19 (11.7%)
Using different outcome definitions	7 (4.3%)
Using different follow-up time	11 (6.8%)
Other (e.g., using different approaches for handling missing data, implementing additional anchored/unanchored comparisons)	12 (7.4%)

*Truong et al. (2023) doi:10.1002/jrsm.1653

Confounding issue: what happens in practice?

A methodological systematic review of studies implementing PAICs*

Limitations acknowledged by authors	Statistics
No acknowledgement	5 (3.1%)
Unmeasured covariates	**136 (84.0%)**
Important covariates not reported in one of the included studies	**60 (37.0%)**
Limited sample size	31 (19.1%)
Heterogeneity across studies	**139 (85.8%)**
Small ESS/little overlap between populations	35 (31.6%)
Lack of a common comparator	23 (14.2%)

*Truong et al. (2023) doi:10.1002/jrsm.1653

Confounding issue: what happens in practice?

B

C

Unanchored PAICs assumptions

- **All** effect modifiers and prognostic variables
  - known
  - adjusted for

- TA592: "**None** of the indirect comparisons provide a **reliable estimate** of relative effectiveness"
- TA567: "the results seemed **implausible**"
- TA540: "**neither** method to be **robust**"
- TA530: "… the concerns about the **robustness** of the simulated treatment comparison"
- TA478: "…**uncertainty** about the **robustness** of the results"
- TA380: "…was **not consistent** with the population in the marketing authorisation"
- …

NICE National Institute for Health and Care Excellence

Quantitative bias analysis (QBA)

QBA: An umbrella term for the methods used to model systematic errors which may distort the results

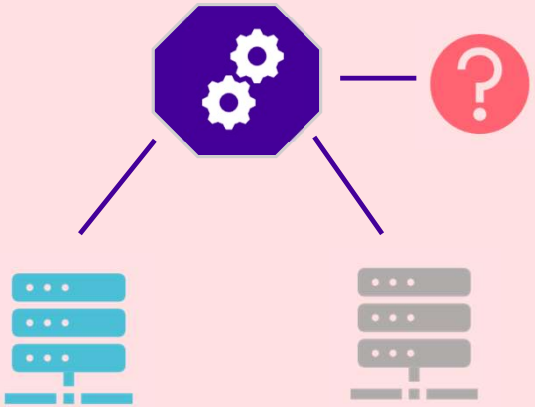
Long history in epidemiology

Aim: To quantitatively measure the direction, magnitude and uncertainty associated with systematic errors on study results

QBA Categorisation		
No unmeasured confounding	Selection, participation and missing data are random within levels of adjusted covariates	No measurement error

Basic idea of QBA

A bias model



Sensitivity/bias parameters
Values need to be pre-specified

Observed data
outcome (Y)
treatment (A)
observed covariates (O)

Unmeasured covariates (U)

Values of sensitivity parameters cannot be estimated from the data alone

Deterministic QBA
Fixed value

Probabilistic QBA
A distribution

Sensitivity analysis for unmeasured confounding for PAICs

Strong assumption: all prognostic factors and effect modifiers are adjusted for

Major concern of unanchored PAICs

In practice, what could be adjusted for depends on data availability

Information on baseline characteristics is limited from the comparator study

QBA for unmeasured confounding via sensitivity analysis

Sensitivity analysis approach based on simulating potential confounder(s)

B — IPD

C — AgD

Study B: IPD

Contains n observations on an outcome Y and $J + L$ observed covariates $\boldsymbol{X} = c(\boldsymbol{O}, \boldsymbol{U})$

Note that $\boldsymbol{U}$ is observed in Study B but not measured in Study C.

$\boldsymbol{X}$

Y	O_1	…	O_J	U_1	⋯	U_L
y_1	$o_{1,1}$	…	$o_{J,1}$	$u_{1,1}$	…	$u_{L,1}$
y_2	$o_{1,2}$	…	$o_{J,2}$	$u_{1,2}$	…	$u_{L,2}$
⋮	⋮	…	⋮	⋮	…	⋮
y_n	$o_{1,n}$	…	$o_{J,n}$	$u_{1,n}$	…	$u_{L,n}$

Study C: aggregate data

Contains reported treatment effect in Study C population $\hat{d}_{C(C)}$, and mean of the marginal distribution for J observed covariates $\boldsymbol{O}$

$E[\boldsymbol{O}] = \overline{\boldsymbol{O}}$

$\hat{d}_{C(C)}$	$E[O_1]$	$E[O_2]$	…	$E[O_J]$

?

$E[U_1] = \tilde{U}_1, E[U_2] = \tilde{U}_2, \ldots, E[U_L] = \tilde{U}_L$

Sensitivity parameters

Deterministic QBA for unanchored STC

C

B

STC: Outcome regression approach*

1. Build regression model based on the IPD from Study B, including all effect modifiers and prognostic factors

$$g\left(\theta_{i(B)}\right) = \beta_0 + \boldsymbol{\beta_1^T X_i}$$

2. Predict the treatment effect for Study C population

$$\hat{d}_{B(C)} = g\left(\hat{\theta}_{B(C)}\right)$$

3. Obtain the unanchored indirect comparison in Study C population, using the prediction from Step 2 and reported aggregate data for Study C

$$\hat{d}_{BC(C)} = \hat{d}_{C(C)} - \hat{d}_{B(C)} = g\left(\bar{\theta}_{C(C)}\right) - g\left(\hat{\theta}_{B(C)}\right)$$

*Ren et al. (2024) doi:10.1002/jrsm.1718

Study C: aggregate data

Contains reported treatment effect for Study C population $\hat{d}_{C(C)}$, and mean of the marginal distribution for J observed covariates $\boldsymbol{O}$

$E[\boldsymbol{O}] = \bar{\boldsymbol{O}}$

$\hat{d}_{C(C)}$ $E[O_1]$ $E[O_2]$ ... $E[O_J]$

Study B: IPD

Contains n observations on an outcome Y and $J + L$ observed covariates $\boldsymbol{X} = c(\boldsymbol{O}, \boldsymbol{U})$

Note that $\boldsymbol{U}$ is observed in the B study but not measured in Study C.

$\boldsymbol{X}$

Y	O_1	...	O_J	U_1	...	U_L
y_1	$o_{1,1}$	...	$o_{J,1}$	$u_{1,1}$	...	$u_{L,1}$
y_2	$o_{1,2}$	...	$o_{J,2}$	$u_{1,2}$	...	$u_{L,2}$
⋮	⋮	...	⋮	⋮	...	⋮
y_n	$o_{1,n}$	...	$o_{J,n}$	$u_{1,n}$	...	$u_{L,n}$

Deterministic QBA

Step 1: Outcome regression model

Build the regression model based on the IPD from Study B

$$g\left(\theta_{i(B)}\right) = \beta_0 + \boldsymbol{\beta_1^T O}_i + \boldsymbol{\beta_2^T U}_i$$

$\boldsymbol{U}$ refers to the unmeasured confounders in Study C and is observed in Study B.

Step 2: Prediction

Predict the treatment effect for Study C population

$$\hat{d}_{B(C)} = g\left(\hat{\theta}_{B(C)}\right)$$

The marginal mean of $\boldsymbol{U}$ for Study C, $\tilde{\boldsymbol{U}} = E[\boldsymbol{U}]$, is the sensitivity parameter.

Assume a fixed value for $\tilde{\boldsymbol{U}}$

$\bar{\boldsymbol{O}}, \Sigma_{IPD}^{O,U}$

NORTA/Gaussian copula to sample individuals

Probabilistic QBA
Specify a distribution for $\tilde{\boldsymbol{U}}$

O_1	U_1
⋮	⋮
O_N	U_N

$$\hat{d}_{B(C)} = g\left(\frac{1}{N}\sum_{i=1}^{N} g^{-1}\left(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^T \boldsymbol{O}_i + \hat{\boldsymbol{\beta}}_2^T \boldsymbol{U}_i\right)\right)$$

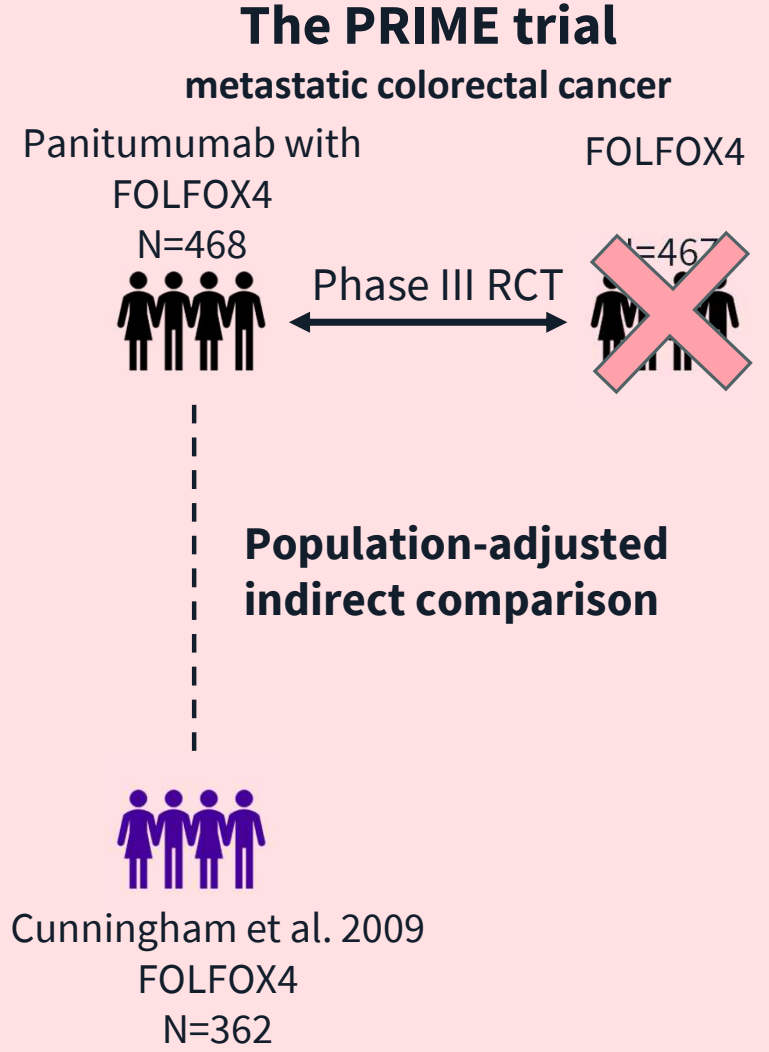
Step 3: Obtain relative effect

Obtain the relative treatment effect using the prediction from Step 2 and reported aggregate data for Study C

$$\hat{d}_{BC(C)} = \hat{d}_{C(C)} - \hat{d}_{B(C)}$$

Case study

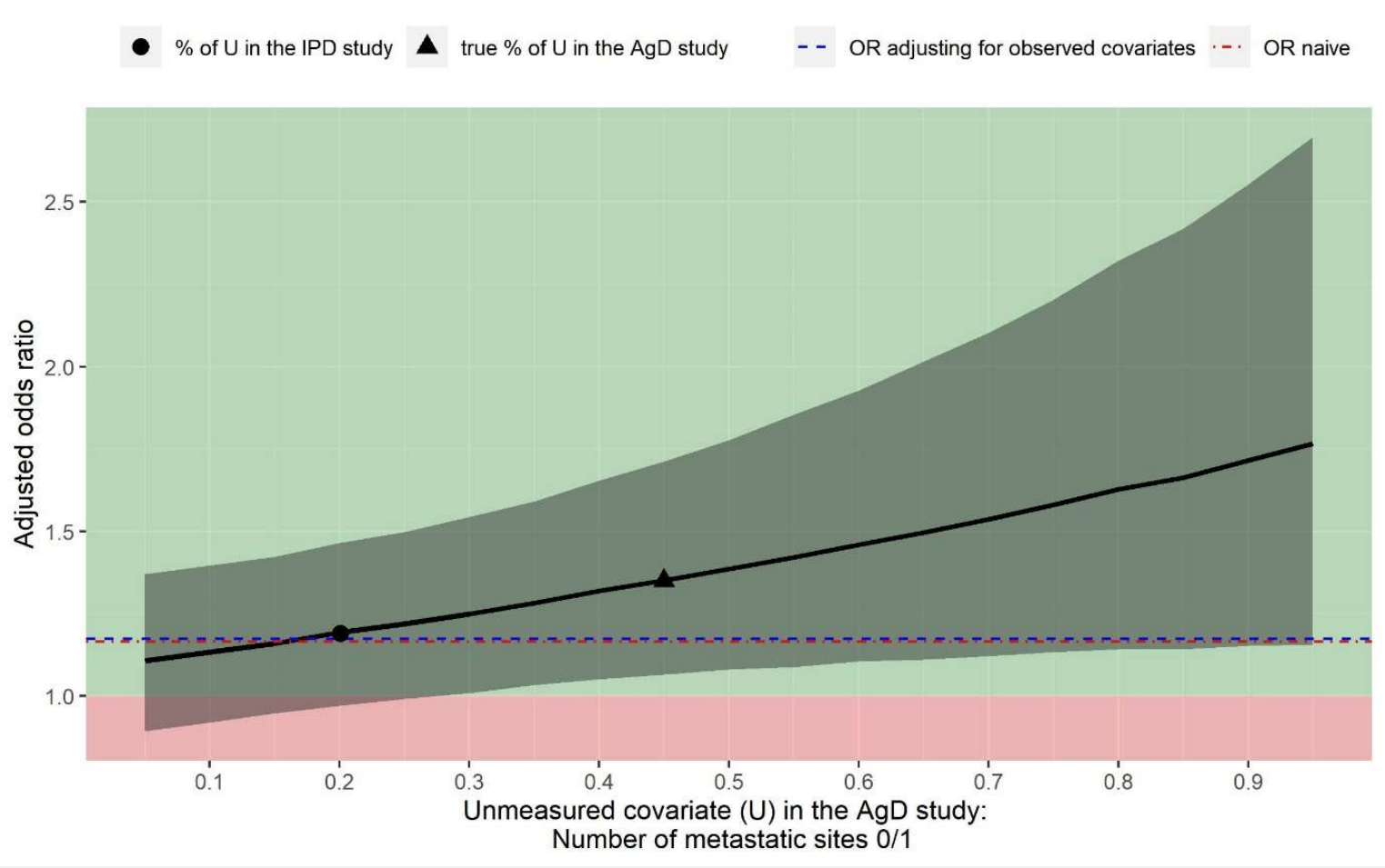
- Re-analyse data from the PRIME trial
- Obtain anonymous IPD for the PRIME trial from the Project Data Sphere® platform
- Objective response rate



Data

	The PRIME trial		Cunningham et al. (2009)
Characteristic	**Panitumumab + FOLFOX4 (n=468)***	**FOLFOX4 (n=467)***	**FOLFOX4 (n=362)**
Male (%)	66	61	65
Age, years (%)			
≤65	60	62	67
65	40	38	33
ECOG performance status (%)			
0/1	95	95	93
≥2	5	5	7
Primary tumour type (%)			
Colon	67	69	56
Rectal and other	33	31	44
Number of metastatic sites (%)			
0/1	20	20	45
≥2	80	80	55
Metastatic site (%)			
Liver alone	18	16	33
Prior adjuvant chemotherapy (%)	15	12	27
Prior surgery (%)	91	91	87
Objective response rate (%)	57.9	53.3	54.1

Sensitivity analysis: number of metastatic sites unmeasured

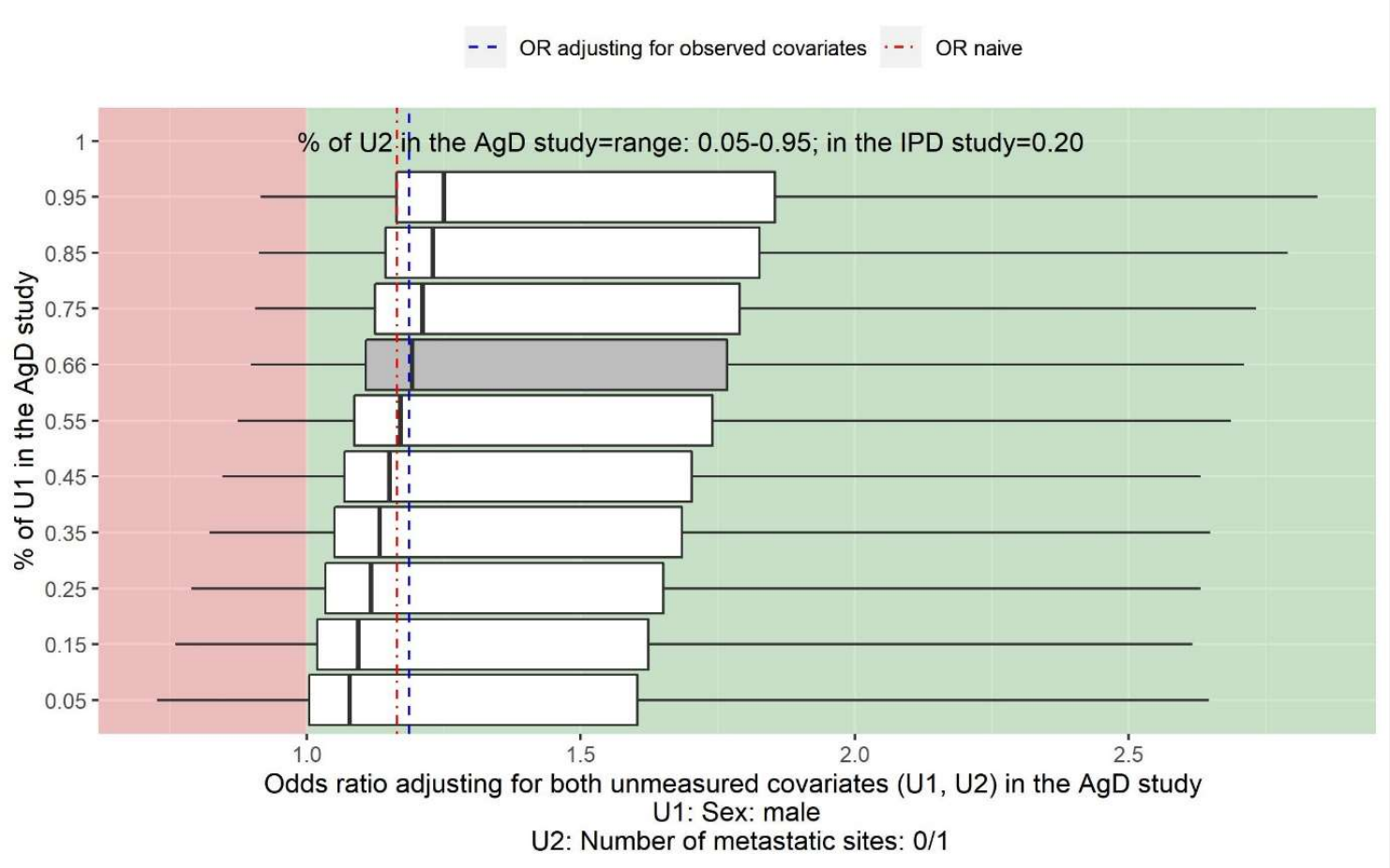


OR from PRIME
1.20 (95% CI, 0.93 to 1.56)

Naïve OR:
1.17 (95% CI, 0.88 to 1.54)

OR adjusted for observed **X**:
1.18 (95% CI, 0.96 to 1.44)

Sensitivity analysis: sex and number of metastatic sites unmeasured



OR from PRIME
1.20 (95% CI, 0.93 to 1.56)

Naïve OR:
1.17 (95% CI, 0.88 to 1.54)

OR adjusted for observed **X**:
1.19 (95% CI, 0.97 to 1.45)

Summary

Unanchored MAIC and STC are **heavily criticised** for its strong assumptions

QBA formally quantifies the bias associated with unmeasured confounding

- Provide a quantitative assessment of the impact of this bias
- Increase the robustness of the ITC approach for single-arm trials

Reference

- Patel D, Grimson F, Mihaylova E, et al. Use of External Comparators for Health Technology Assessment Submissions Based on Single-Arm Trials. Value Health. Aug 2021;24(8):1118-1125. doi:10.1016/j.jval.2021.01.015
- Truong B, Tran LT, Le TA, Pham TT, Vo TT. Population adjusted-indirect comparisons in health technology assessment: A methodological systematic review. Res Synth Methods. Sep 2023;14(5):660-670. doi:10.1002/jrsm.1653
- Cunningham D, Sirohi B, Pluzanska A, et al. Two different first-line 5-fluorouracil regimens with or without oxaliplatin in patients with metastatic colorectal cancer. *Ann Oncol*. Feb 2009;20(2):244-50. doi:10.1093/annonc/mdn638
- Ren S, Ren S, Welton NJ, Strong M. Advancing unanchored simulated treatment comparisons: A novel implementation and simulation study. Res Synth Methods. 2024 Apr 8. doi: 10.1002/jrsm.1718. Epub ahead of print. PMID: 38590103.

Challenges associated with external control arms drawn from the real world for deriving relative effectiveness for HTA purposes: a UK case study

PSI Conference 2024 session: “Bias in indirect treatment comparisons and evolving methodology: implications for health technology assessment and beyond”

Nicolas Scheuer, Health Outcomes Partner, Roche Products Ltd, United Kingdom

18 June 2024

Disclaimer

Views or opinions expressed in this presentation are solely my own and do not necessarily express the views or opinions of Roche.

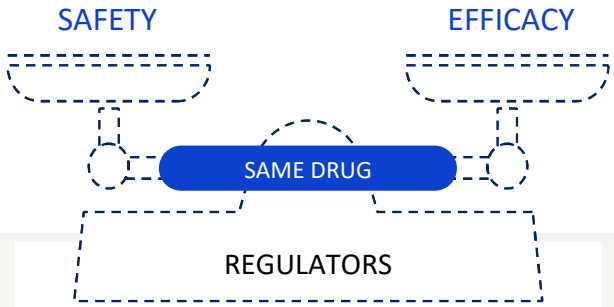
Flatiron Health is an independent affiliate of the Roche Group.

Outline

- Regulators vs payers' remit
- HTA bodies in the UK
- Decision uncertainty & the premise of RWE
- Pralsetinib case study
- NICE & SMC recommendations
- Concluding thoughts

Questions asked by HTA agencies/payers are different to those asked by regulators and RWE can help answer these questions

Does efficacy outweigh safety?

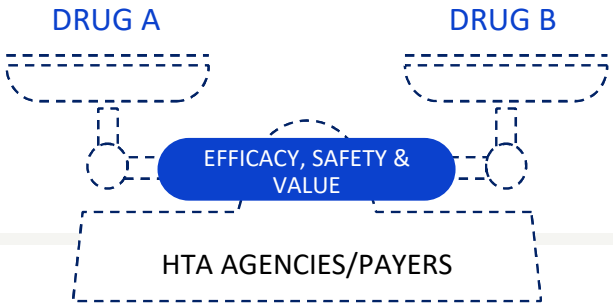


Positive benefit-risk of a new drug

More sophisticated packages of evidence are required to justify value, price and reimbursement to payer decision-makers

RWE is an essential element of that additional evidence packages used to justify value, price and reimbursement

Does the product demonstrate added value over current treatment?

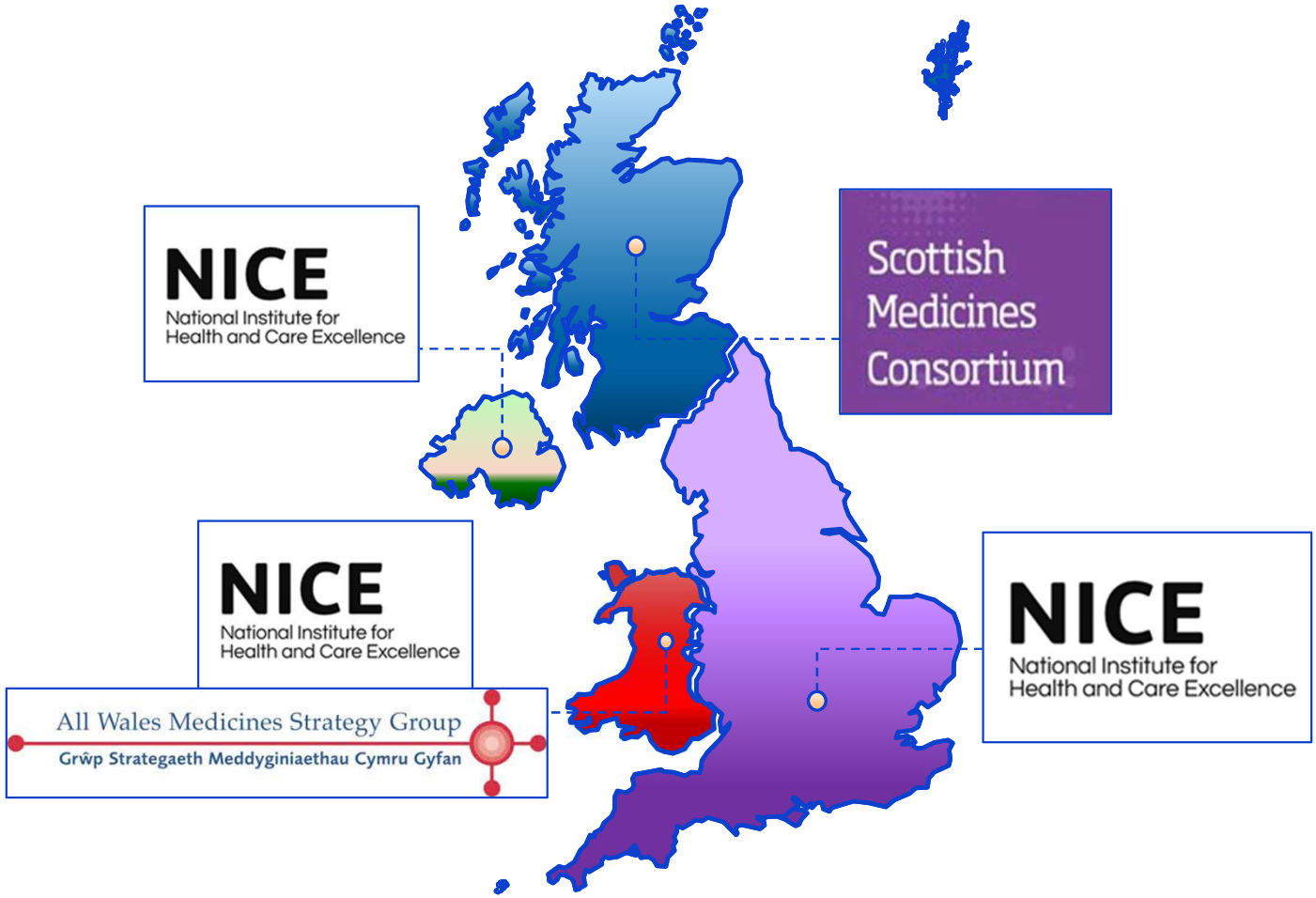


Unmet need	Is the product needed?
Clinical effectiveness	Does it work?
Safety	Is it safe?
Restrictions	In which population does it best work?
Comparative effectiveness	How well does it work vs SoC?
QoL	Does it improve a patient's QoL?
Cost-effectiveness	Is it worth it?
Opportunity cost	What investments are to be displaced?
Budget impact	Can we afford it?
Patients perspective	What is the patient's view
Political imperative	Do we need to fund it?

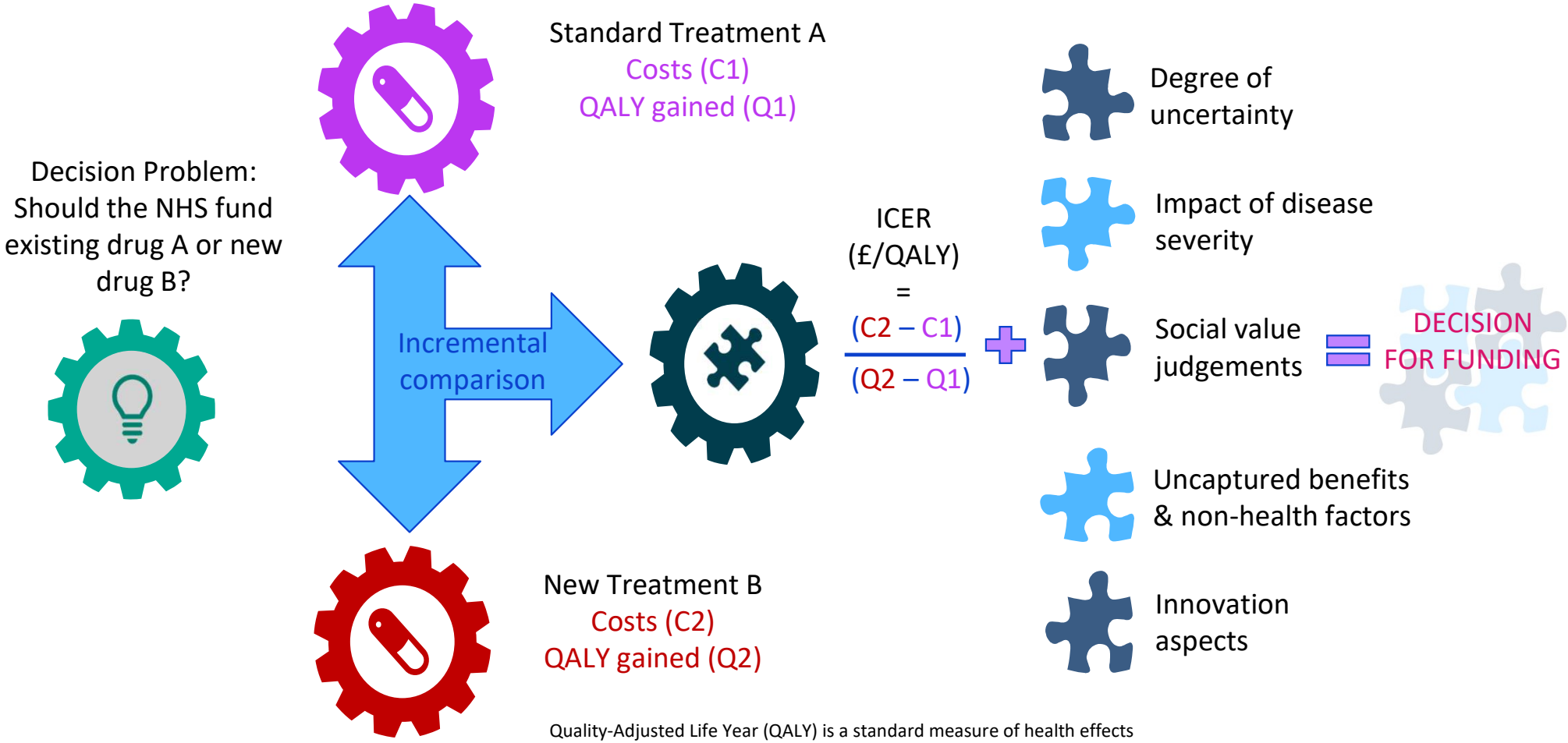
HTA: health technology assessment, QoL: quality of life, SoC: standard of care, RWE: real world evidence

Adapted from EUnetHTA (https://www.eunethta.eu/wp-content/uploads/2018/03/HTACoreModel3.0-1.pdf (last accessed 24 May 2024)

HTA bodies in the UK



How manufacturer evidence is evaluated by payers in the UK



ICER: Incremental Cost-Effectiveness Ratio
NHS: National Health Service

Quality-Adjusted Life Year (QALY) is a standard measure of health effects that combines length of life (survival) with quality of life
Typical ICER threshold is £20-30k per QALY

Decision-making is both uncertain & complex

Small sample

Limited evidence potentially increases the (costly) risk of making the wrong decision

Evidence gaps

Eg. in Population, Intervention, Comparator, Outcome, Study design (PICOS), extrapolations to a lifetime horizon

Decision making

Generalisability

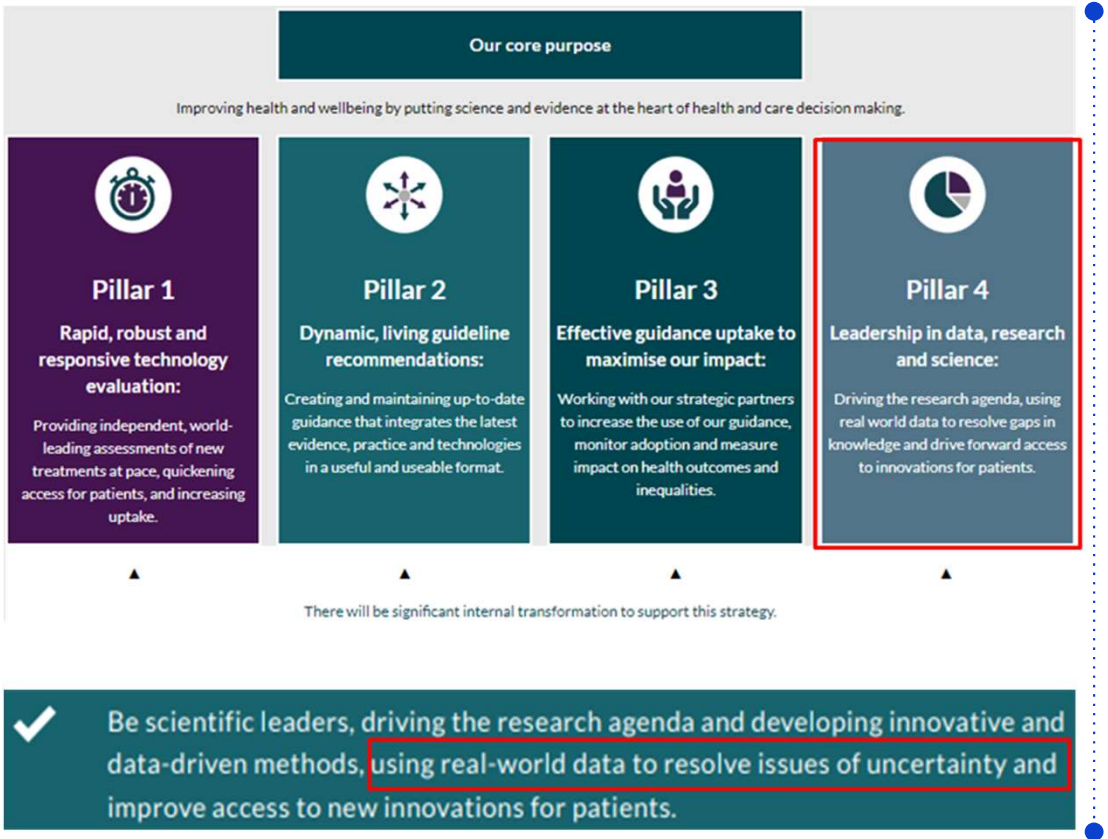
Internal vs external validity, or transportability to another setting

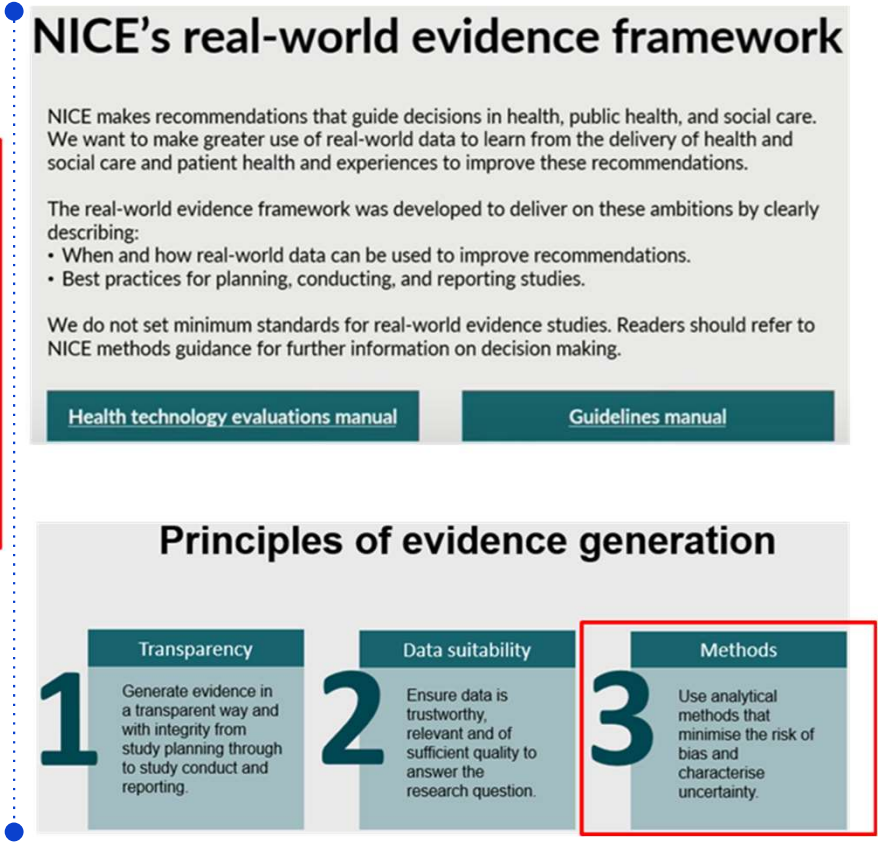
Heterogeneous evidence sources

Decisions should be informed for the right type of patients

Adapted from "Decision Modelling for Health Economic Evaluation - Foundations Course" (2022), University of York

The bigger picture: RWE to play a critical role in resolving NICE's decision uncertainties & evidence gaps to drive forward access to innovations for patients





NICE 2021-2026 strategy & RWE framework 2022

Key information at time of dossier submission (mid 2021)

Indication: Pralsetinib for RET fusion-positive advanced NSCLC (RET+ aNSCLC for simplicity)

Clinical evidence: ARROW trial

Small patient number

- In population: ~1-2% of NSCLC
- In trial: 116 (untreated), 165 (pre-treated)

Unusual clinical presentation

Patients tend to be younger, non-smokers and female compared to their wild type counterparts

Immature efficacy

Clinical efficacy was promising, yet highly uncertain due to low patients number, immature follow-up data & absence of comparator (single arm)

Key analysis considerations

Treatment comparators
(in the untreated population)

- Pembrolizumab plus pemetrexed and chemotherapy (the focus here)
- Platinum-based chemotherapy with or without pemetrexed

Real-world data source to inform the indirect treatment comparison

Flatiron Health enhanced data-mart, under the assumption that RET fusion status is not prognostic

Analysis method & estimand

- IPTW was used to adjust for patient differences
- The chosen estimand was the ATT, i.e., in an ARROW-like population
- Quantitative bias analysis to quantify uncertainty & residual bias

Gainor et al. (2021). *Pralsetinib for RET fusion-positive non-small-cell lung cancer (ARROW): a multi-cohort, open-label, phase 1/2 study*

aNSCLC: advanced non-small-cell lung cancer, ATT: average treatment effect among the treated, IPTW: Inverse probability of treatment weighting, ITC: indirect treatment comparison, RET: rearranged during transfection

Baseline characteristics of ARROW patients and Flatiron EDM cohort given 1st line pembrolizumab with chemotherapy

Before (left) and after (right) IPTW adjustment; SMD<0.1 indicates sufficient balance

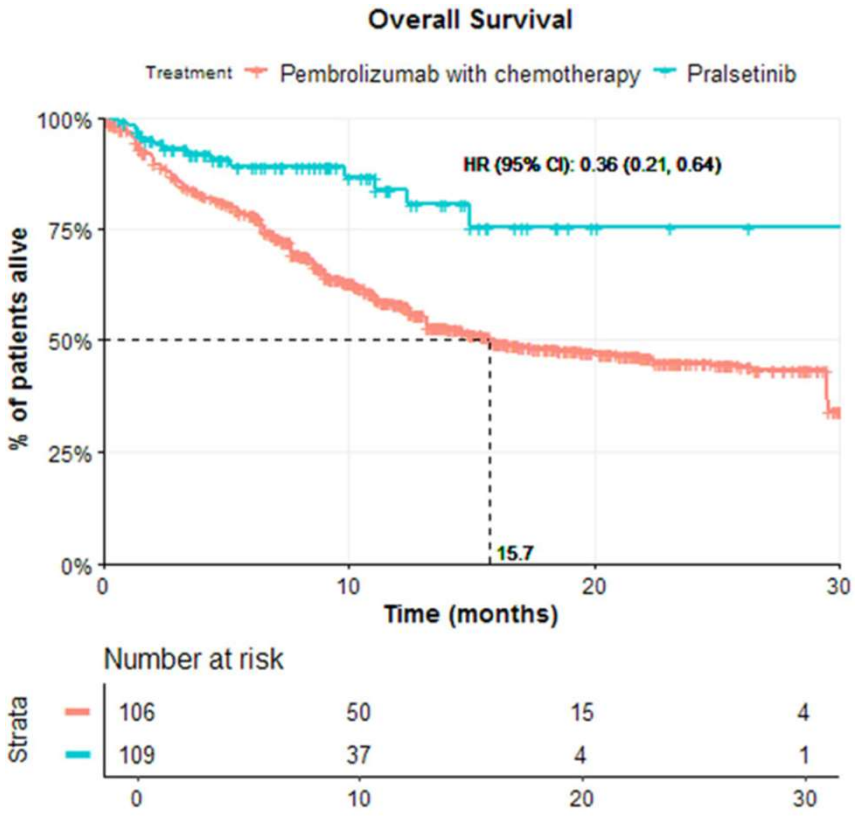
	Level	Pembrolizumab with chemotherapy	Pralsetinib	SMD	Pembrolizumab with chemotherapy	Pralsetinib	SMD	Adjusted
N		1270	109		217/1270	109/109		
Age (%)	<65	508 (40.0)	65 (59.6)	0.4	58.9	59.6	0.015	Y
	>=65	762 (60.0)	44 (40.4)		41.1	40.4		
Sex (%)	F	569 (44.8)	59 (54.1)	0.187	54.5	54.1	0.007	Y
	M	701 (55.2)	50 (45.9)		45.5	45.9		
Smoking history at baseline (%)	History of smoking	1144 (90.1)	43 (39.4)	1.25	40.3	39.4	0.017	Y
	No history of smoking	126 (9.9)	66 (60.6)		59.7	60.6		
ECOG (%)	0	512 (40.3)	34 (31.2)	0.191	32.9	31.2	0.037	Y
	1	758 (59.7)	75 (68.8)		67.1	68.8		
Time from initial diagnosis to first dose (months) (median [IQR])		1.18 [0.76, 1.84]	1.74 [1.25, 2.30]	0.148	1.32 [0.92, 2.24]	1.74 [1.25, 2.30]	0.042	Y
Stage at initial diagnosis (%)	STAGE I, II, or III	204 (16.1)	17 (15.6)	0.013	16.6	15.6	0.028	Y
	STAGE IV	1066 (83.9)	92 (84.4)		83.4	84.4		
Race (%)	White	883 (69.5)	54 (49.5)	0.573	52.3	49.5	0.061	Y
	Other	248 (19.5)	49 (45.0)		41.9	45		
	Unknown	139 (10.9)	6 (5.5)		5.8	5.5		
Brain/CNS metastasis only (%)	0	1090 (85.8)	79 (72.5)	0.333	87.5	72.5	0.383	N
	1	180 (14.2)	30 (27.5)		12.5	27.5		

Effective sample size / N

CNS: central nervous system, ECOG: Eastern Cooperative Oncology Group, EDM: enhanced data-mart, IPTW: inverse probability of treatment weighting, IQR: interquartile range, SMD: standardised mean difference

Adapted from Popat et al (2022). *Addressing challenges with real-world synthetic control arms to demonstrate the comparative effectiveness of Pralsetinib in non-small cell lung Cancer*

NICE committee members dismissed the Flatiron-informed relative effect estimate due to challenges in assessing its quality



Popat et al (2022)

"The committee expressed concerns about the appropriateness of the real-world data in the Flatiron database, due to the challenges in assessing its quality."

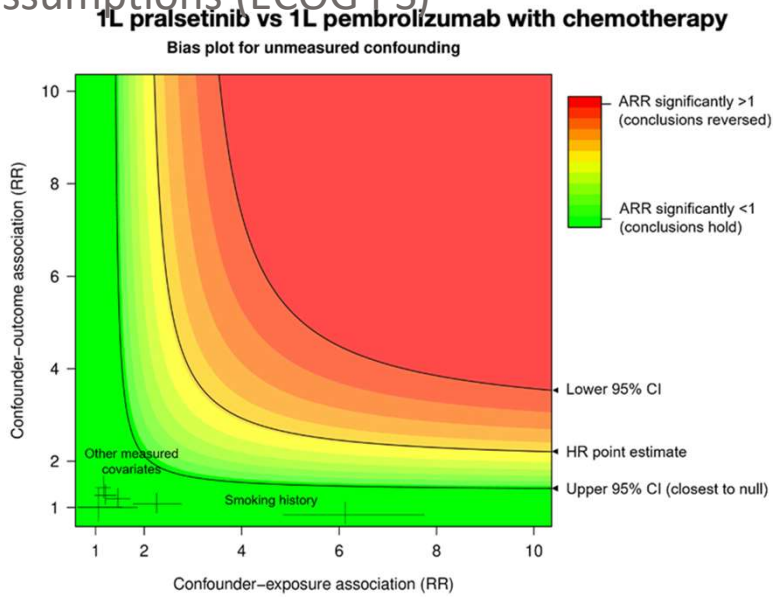
"It also noted that an indirect treatment comparison of clinical trial data with real-world data can be expected to introduce bias because the care that people have in each setting is likely to be different."

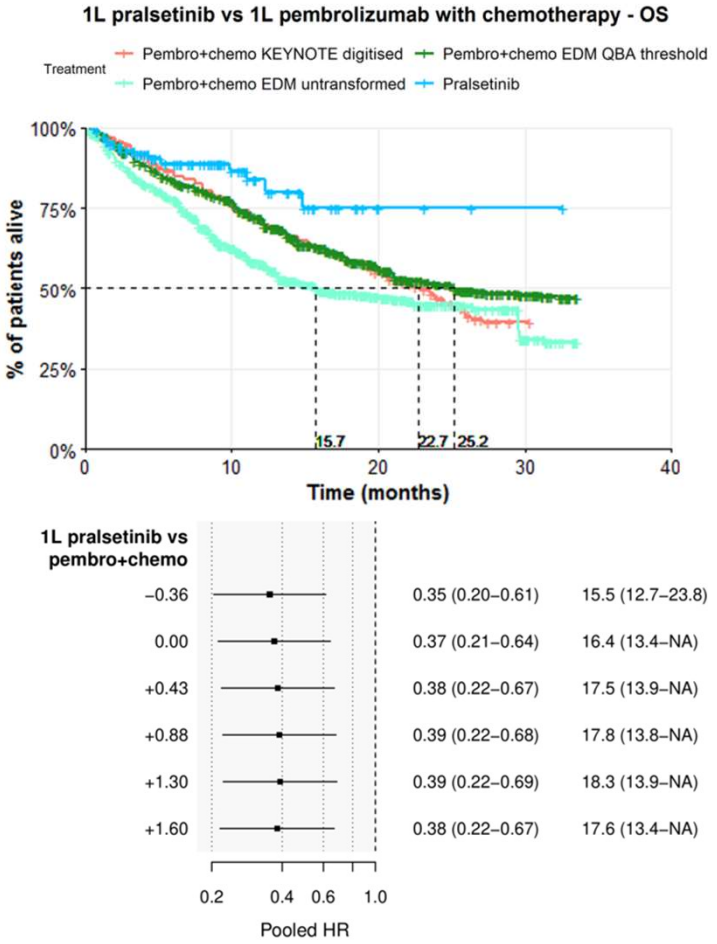
"For these reasons, the hazard ratio results of the indirect treatment comparison may have overestimated the relative clinical effectiveness of pralsetinib."

Pralsetinib [TA812], NICE Appraisal Consultation Document (2022)

This decision was reached despite efforts to characterise the uncertainty and risk of residual bias through QBA

QBA for unmeasured confounding, poorer real-world performance and missing data assumptions (ECOG PS)





QBA sensitivity analyses suggest that results are robust to plausible unmeasured confounding, extreme deviations from random missingness for baseline ECOG PS, and non-conformance of treatment performance in the real-world vs clinical trial setting

Popat et al (2022) & Roche data on file

ECOG PS: Eastern Collaborative Oncology Performance Status, QBA: quantitative bias analysis

SMC was more amenable to trial-RW informed evidence (vs NICE) leading to a positive recommendation

NICE

NEGATIVE decision (Jun 2022)

Contributing factors

- NICE did not show increased acceptance of *either* evidence uncertainty for a rare disease *or* RWE to address data gaps – two aspects that the 2022 methods review stated to address*
  - That said, Flatiron analysis was not pre-specified, nor published on a publicly accessible platform (e.g., the Real-World Evidence Registry)
- No EAG critique of Flatiron comparison – issues "unearthed" during 1st committee meeting
  - Introduction of the PDC comparator in the untreated pop
  - Naive comparison for pralsetinib vs pembrolizumab + PDC
- Challenging for precision medicines to demonstrate cost-effectiveness vs high-uptake drugs such as pembrolizumab
- Not reaching "end of life" criteria in the untreated pop

* However, pralsetinib submission was assessed under the 2013 NICE methods guide

SMC

POSITIVE decision (Mar 2023)
Interim funding & subject to reassessment in both untreated & pre-treated populations

Contributing factors

- Less relevant comparators of interest
- More receptive to Flatiron comparative analysis
  - Multiple opportunities to emphasise its robustness
- Reaching "end of life" criteria in both populations
- Cost-effectiveness improving factors
  - More resource use in general
  - Increased cost of intravenous

Key take-away

- Generate payer-grade evidence by following existing guidances, e.g., NICE RWE Framework (2022)
- Ensure reproducible, real-world data quality
  - Data fitness for purpose (relevance & reliability)
  - Data provenance (curation, governance)
  - Research transparency (integrity, methods)

EAG: External Assessment Group, NICE: National Institute for Health and Care Excellence, PDC: Platinum-based doublet chemotherapy, RW: real world, SMC: Scottish Medicines Consortium

Some concluding thoughts

1. It is about time to reimbursement
   Especially in two-tier healthcare systems

2. Externally controlled single-arm trials are increasingly being submitted for consideration to regulators & payers
   Driven by rare patient populations, ethical considerations, rapidly evolving standard of care, complex technologies, drug development costs

3. Real-world evidence to take a prominent role in filling evidence gaps - to be aligned with relevant guidance to provide trustworthy, quality, reproducible & transparent data to decision-makers
   This is essential for comparative effectiveness for use in economic evaluation, where treatment effect is a key decision driver

4. Build strategic partnerships with key stakeholders (e.g., academia, HTA bodies) to advance scientific knowledge
   Keep pushing uphill: what is challenging today might become mainstream tomorrow (think population-adjusted methods)

Thank you for your attention

Doing now what patients need next

EFSPI EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

Panel Discussion

"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."

Slide 72

1 Do we want to have slides with pre-determined questions displayed? Or do we just leave this header slide up during the panel and not display any questions asked verbally?
PSI HTA SIG, 12/05/2023

EFSPI EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

Closing Remarks

"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."

HTA ESIG

Want to get more involved in this discussion and help impact the future of EU HTA? Become a member of the HTA ESIG today – scan the QR code or email htasig@psiweb.org

You can also join us for our social catch-up during Wednesday lunch break - all welcome!

Finally, don't miss the HTA townhall closing out the conference again this year!

Visit Us

"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."