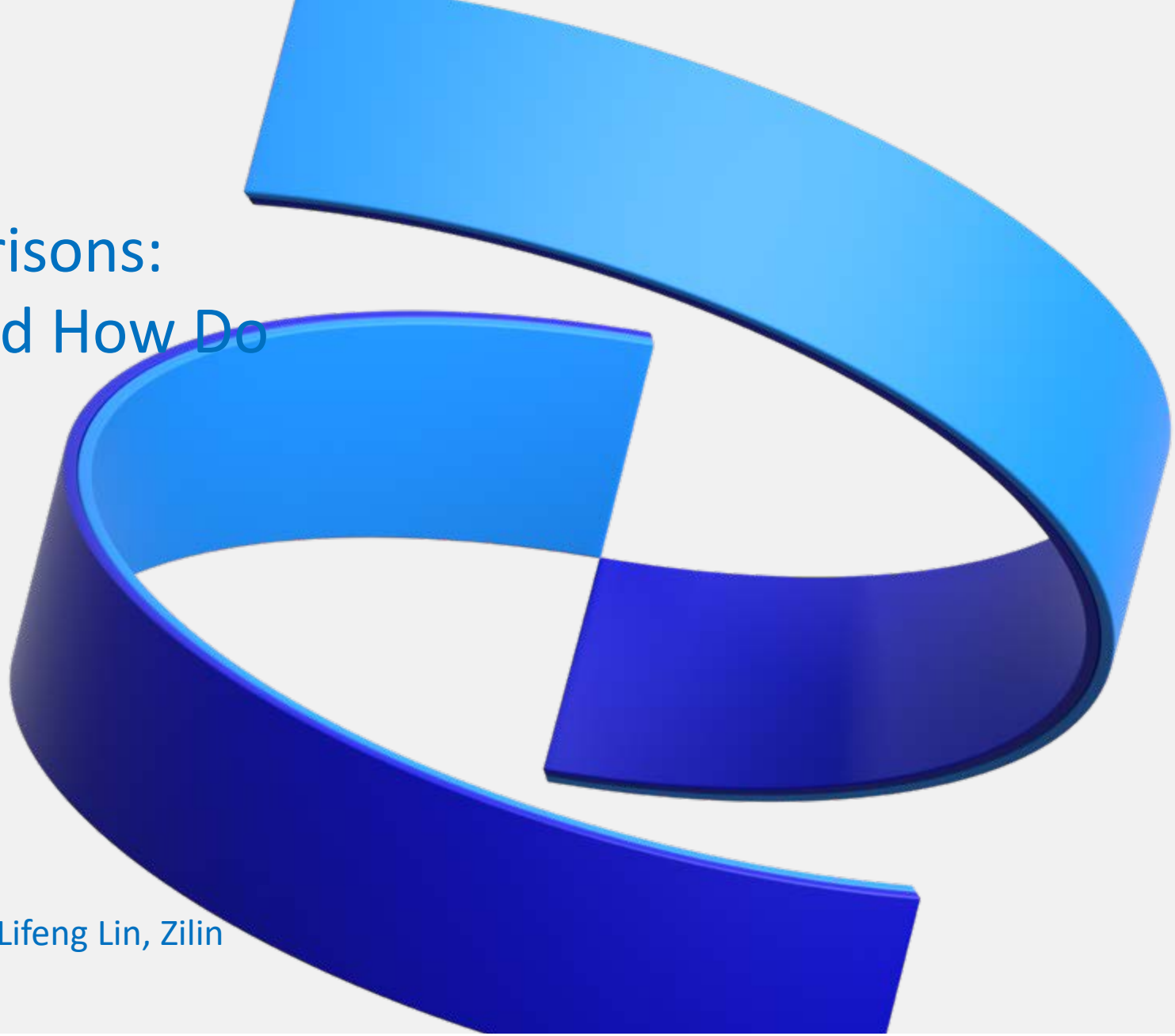
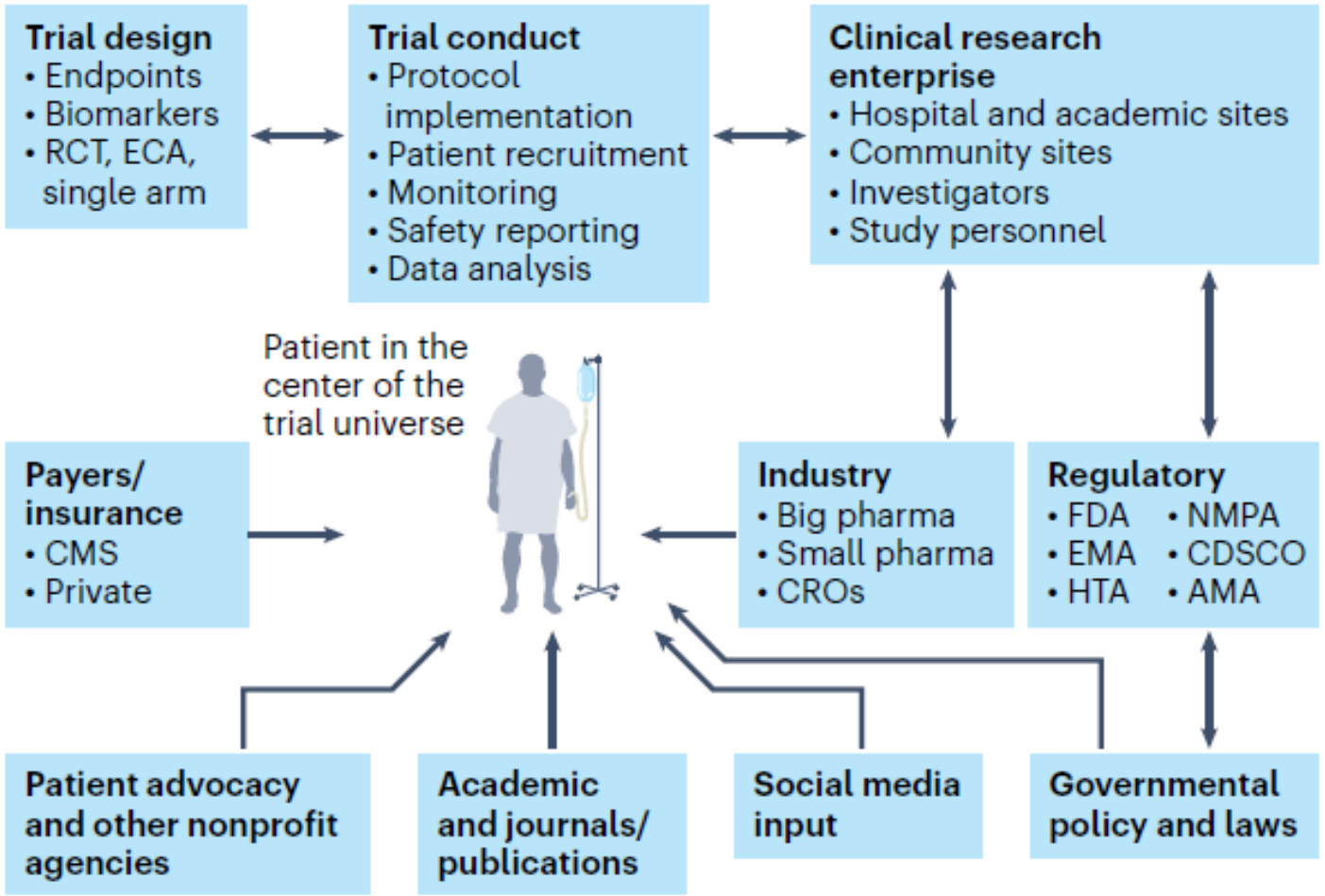
Indirect Treatment Comparisons: When Are They Needed and How Do They Work?



Haitao Chu, MD, PhD, FASA

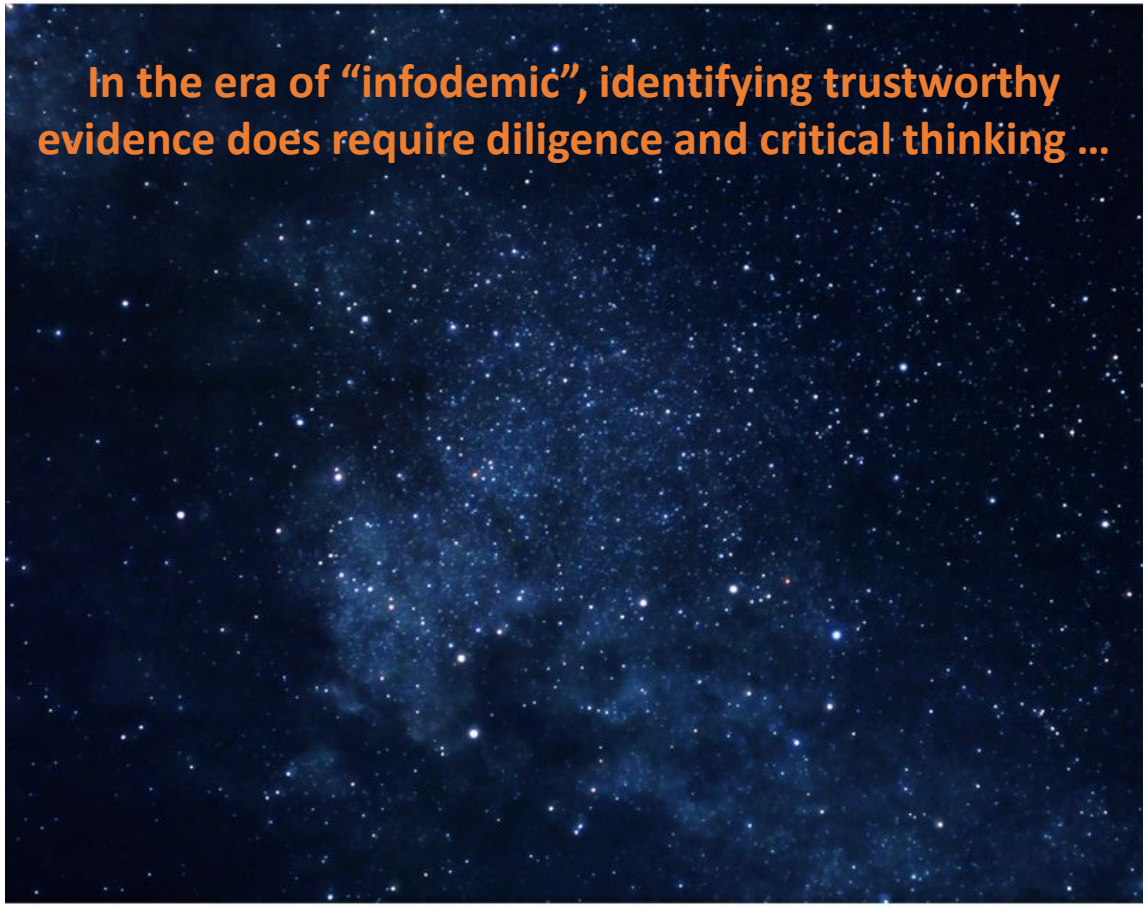
ASA-CT, April 14 2023
Groton, CT

In collaboration with Zheng Wang, Yong Chen, Lifeng Lin, Zilin Wang and Joseph C Cappelleri



The patient as the center of the clinical trial universe in the clinical research enterprise. The main constituents of the clinical trial enterprise — patients, academic centers, industry sponsors (big and small pharma), government/cooperative group sponsors, regulatory agencies, patient advocacy organizations and CROs—need to work together, with the patient as the center of this clinical trial universe. AMA, African Medicines Agency; CDSCO, Central Drugs Standard Control Organization (India); CMS, Centers for Medicare and Medicaid Services; ECA, external control arm; EMA, European Medicines Agency; HTA, Health Technology Assessment; NMPA, National Medical Products Administration (China).

Subbiah V. The next generation of evidence-based medicine. Nature Medicine 2023, 49-58.



Night Sky Constellations

The Big Dipper

ITC plays a Crucial Role in Health Technology Assessment

- When considering a new technology for reimbursement, private and national payers prefer evidence from direct treatment comparisons by standard pairwise meta-analysis of head-to-head randomized controlled trials (RCTs) comparing the new technology to the current established practice or standard of care in the patient population for which the new technology is indicated.

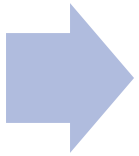
- In the absence of head-to-head RCTs meeting these requirements, payers often expect to see evidence from indirect treatment comparisons including Bucher's method, network meta-analysis (NMA) or population adjusted indirect comparison (PAIC) to demonstrate the clinical value of the new technology.

Health Technology Assessment (HTA)

- The HTA dossier is typically submitted to a national agency, such as The National Institute for Health and Care Excellence (NICE) in England, who assesses the product's clinical and economic value relative to current clinical practice.
- The HTA agency determines whether a product is deemed to provide sufficient incremental value at an acceptable price to justify its use by the health service.
- A successful HTA submission is one of the most significant hurdles in the market access journey.

The feasibility assessment of performing an ITC (NMA vs. PAIC):

Is there a **connected network** (via common comparators) to compare the treatments of interest regarding the outcomes of interest?
NO



Are there **minimal differences** in study design, patient, or outcome characteristics across comparisons that are likely modifiers of the relative treatment effects?
NO

PAIC methods may be considered since they do not require a connected network or common comparator and can adjust for differences in the baseline risk factors between the treatment groups.

Bucher’s Method for Indirect Comparisons (only aggregate data)

Bucher's method for indirect comparisons

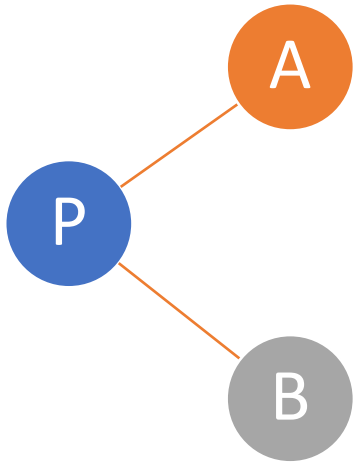


Fig 1A

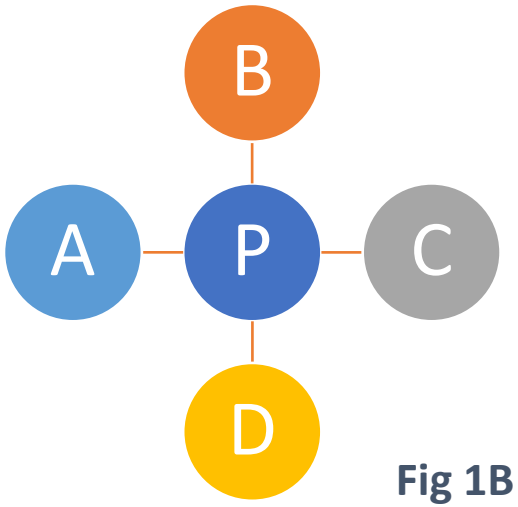


Fig 1B

- Can be used in a simple ITC to compare outcomes between A and B (Fig 1A) or across a star-shaped network of treatments, where several treatments are compared to common comparator P (Fig 1B).
- Assume that the trials included in the ITC are similar with regards to the study design, population, outcome measurements, and the distribution of treatment effect-modifiers (i.e., study and patient characteristics that have an independent impact on treatment outcome); and relative effects are transportable;
- **Unsuitable for performing ITCs with more complex networks of treatments**, e. g. closed loops or multi-arm trials.

Bucher's method for adjusted indirect comparisons

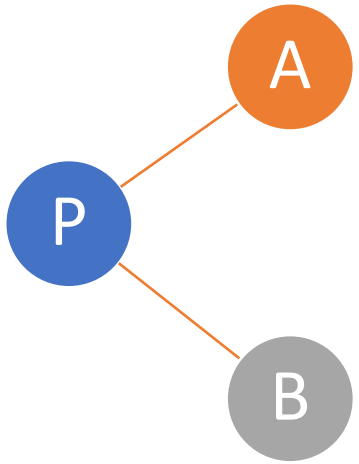


Fig 1A

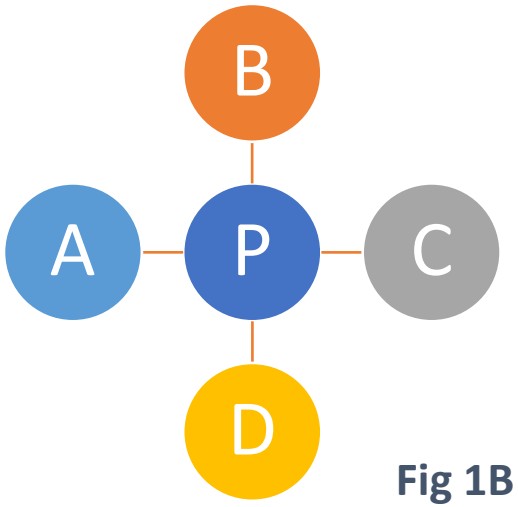


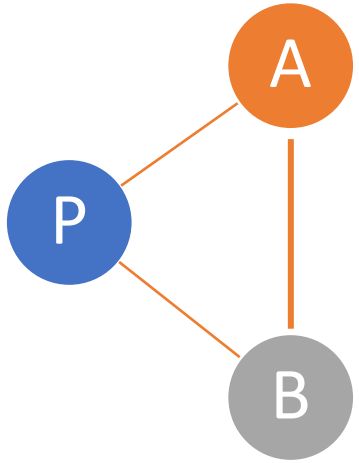
Fig 1B

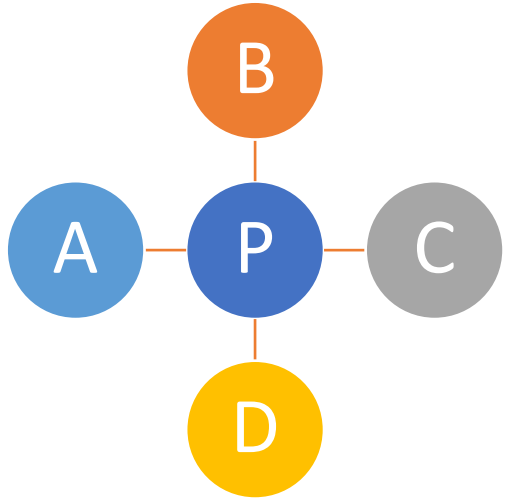
- Focus on relative effects to keep randomization
- Compute relative effect of **A vs P,** such as log risk ratio **$\log RR_{AP}$** with corresponding **SE_{AP}**, and relative effect comparing **B vs P** (e.g. **$\log RR_{BP}$** and **SE_{BP}**),
- The indirect comparison effect estimate of **A vs B** will be the difference between the two relative effect estimates (i.e. **$\log RR_{AB}$** will be computed as $\log RR_{AB} = \log RR_{AP} - \log RR_{BP}$ with $SE_{AB} = \sqrt{SE_{AP}^2 + SE_{BP}^2}$)
- Transitivity in one measurement scale means non-transitivity in another scale unless under the null

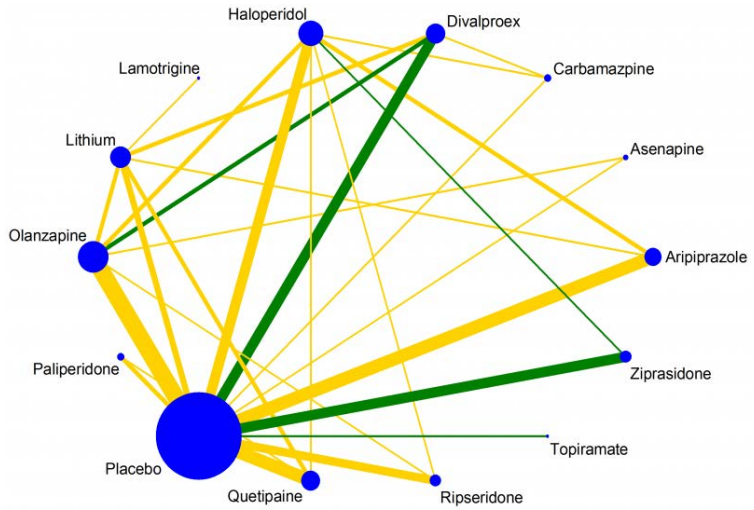
Network Meta-analysis (NMA)
(only aggregate data)

Network Meta-analysis (NMA)

NMA expands the scope of a conventional pairwise meta-analysis to simultaneously compare multiple treatments, synthesizing both direct evidence within randomized controlled trials (RCTs) and indirect evidence across RCTs to improve statistical precision and reduce bias.







Network Meta-analysis (NMA): Key Points

- Network meta-analysis is a technique for comparing three or more interventions simultaneously in a single analysis by combining both direct and indirect evidence across a network of studies.
- Network meta-analysis produces estimates of the relative effects between any pair of interventions in the network, and usually yields more precise estimates than a single direct or indirect estimate. It also allows estimation of the ranking and hierarchy of interventions.
- A valid network meta-analysis relies on the assumption that the different sets of studies included in the analysis are similar, on average, in all important factors that may affect the relative effects.

Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022.

Network Meta-analysis (NMA): Key Points

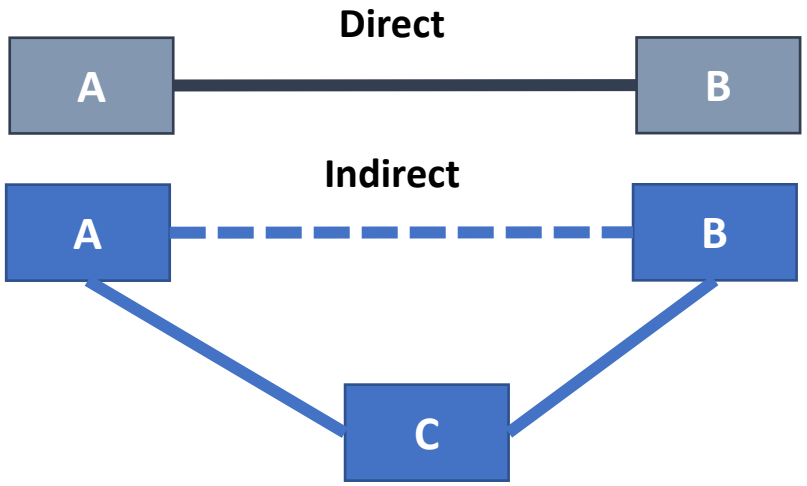
- Incoherence (also called inconsistency) occurs when different sources of information (e. g. direct and indirect) about a particular intervention comparison disagree.
- Grading confidence in evidence from a network meta-analysis begins by evaluating confidence in each direct comparison. Domain-specific assessments are combined to determine the overall confidence in the evidence.

Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022.

Network Meta-analysis (NMA)

In addition to the assumptions that the trials included in the ITC are similar and relative effects are transportable (CB-NMA), NMA typically assumes consistency between direct and indirect evidence.

Effect: A vs B (Direct) **? = ?** **A vs C – B vs C (Indirect)**

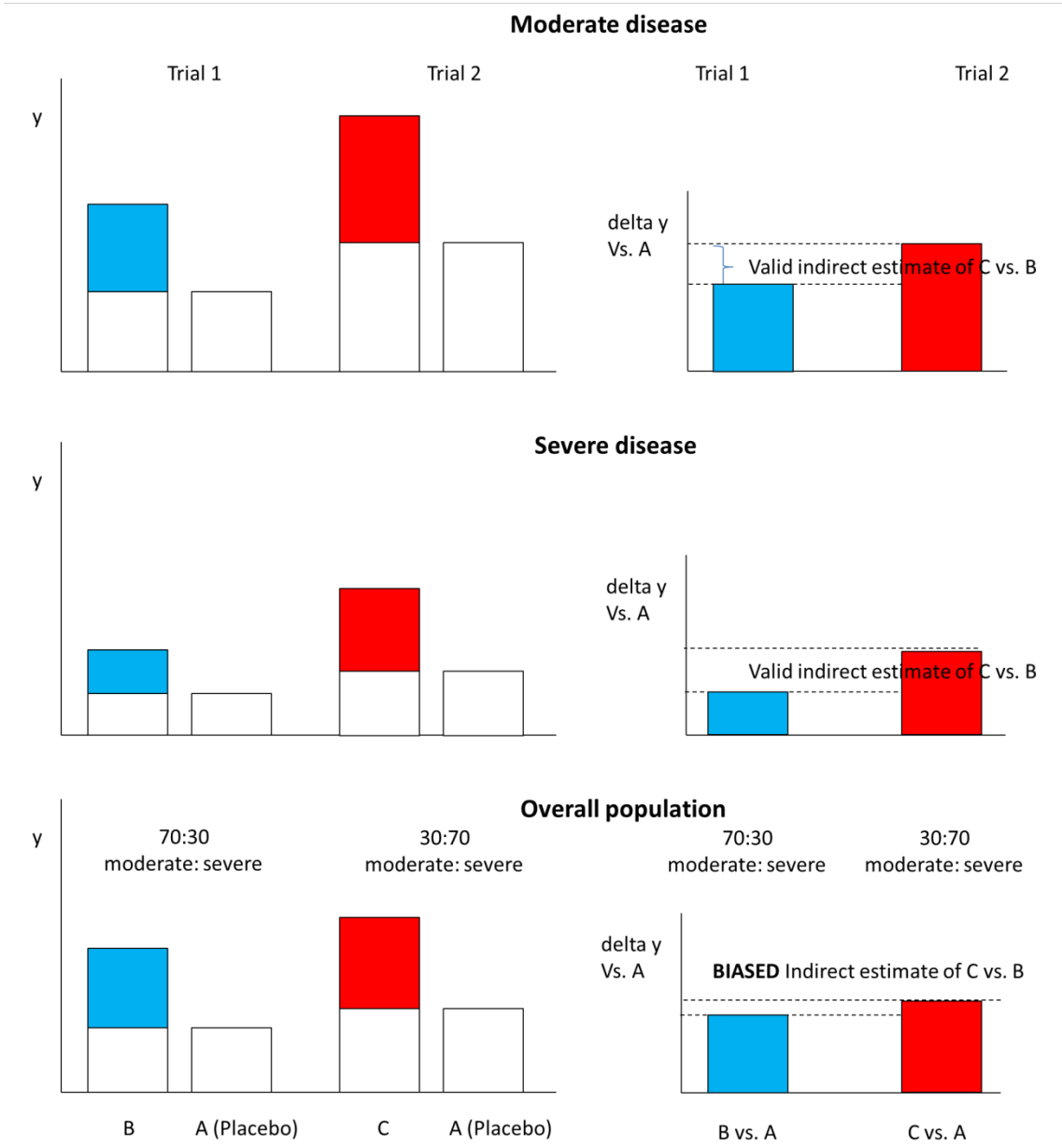


- **Consistency (Transitivity) Equation**

  $\delta_{AB} = \delta_{AC} - \delta_{BC}$ **for** $A \neq B \neq C$

- **Choice of Effect Measures Matters**

Valid and biased indirect comparisons: Effect modification

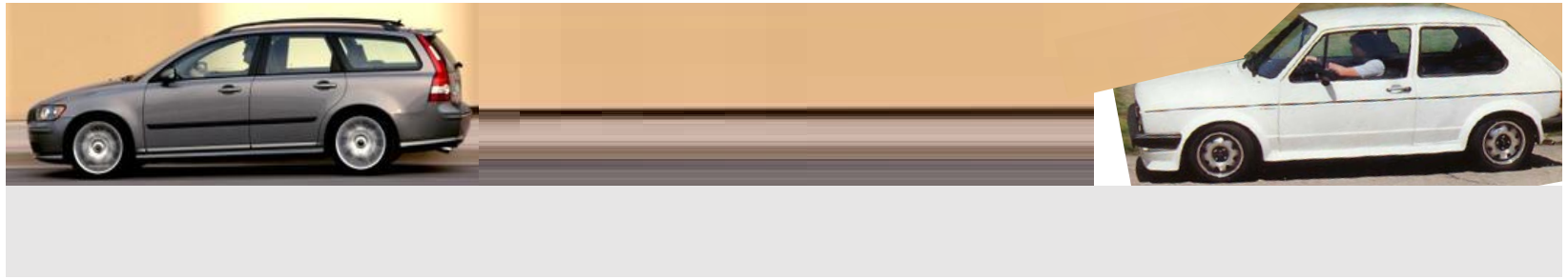


‘Trial 1: Porsche versus Golf’

Porsche - Golf = 2s

‘Trial 2: Volvo versus Golf’

Volvo - Golf = 8s

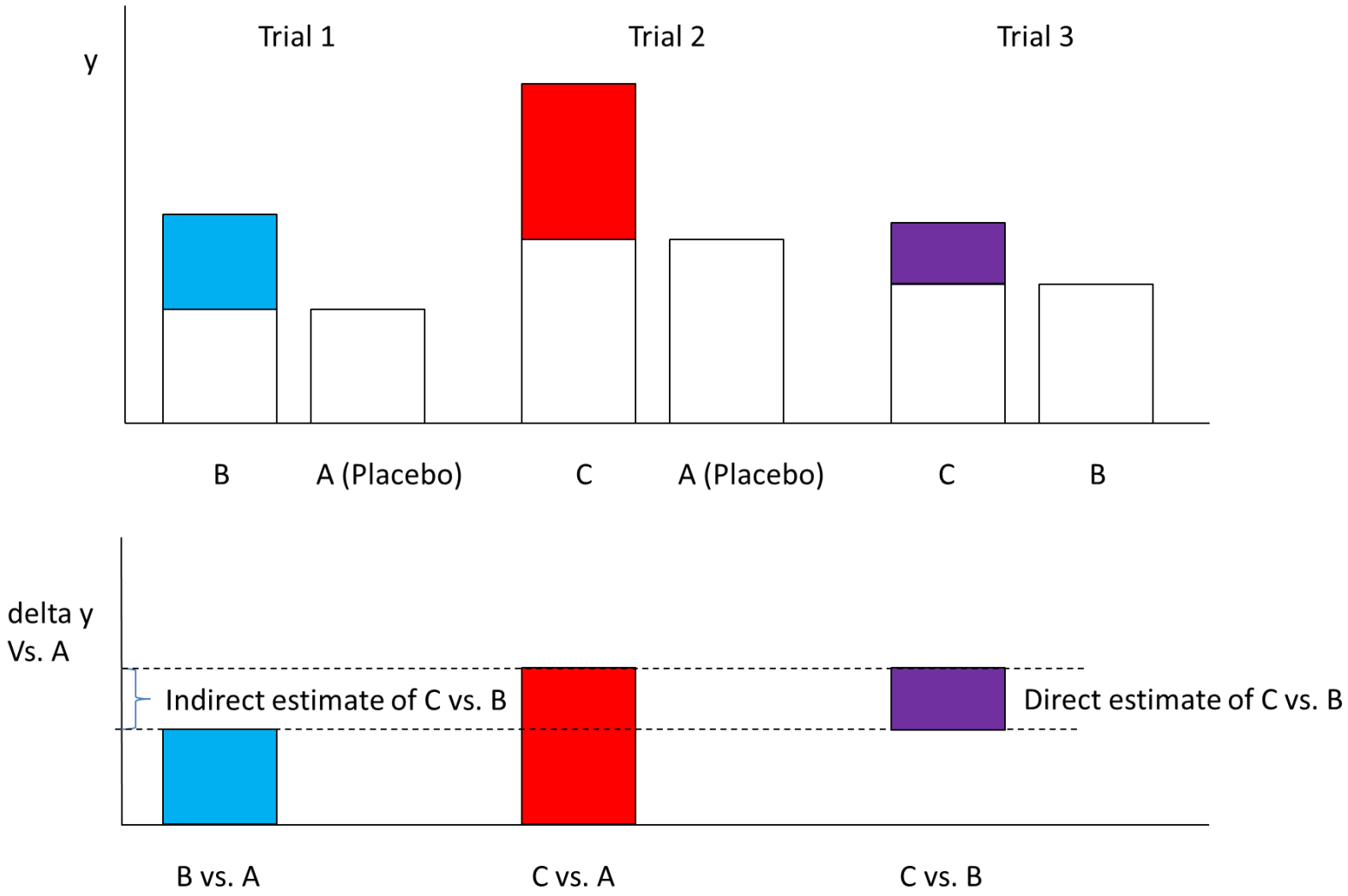


⟶ Volvo versus Porsche: 8-2=6s (Indirect comparison)

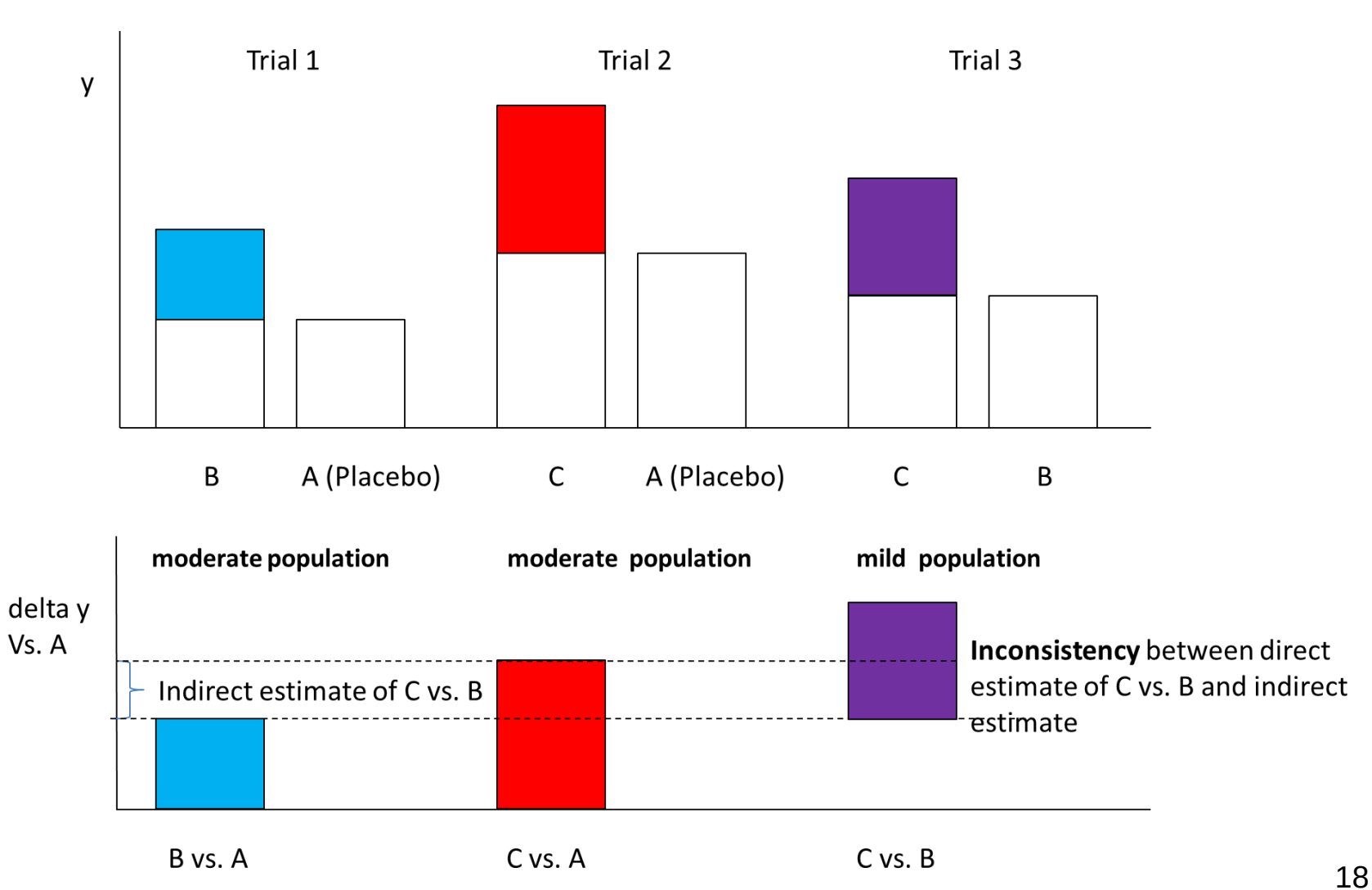
Is a Volvo faster than a Porsche?

No, biased indirect estimate due to imbalance in treatment effect modifier (snow) across comparisons

Consistency in NMA



Inconsistency in NMA



18

Network Meta-analysis (NMA):

Contrast-Based (CB) NMA vs. Arm-Based (AB) NMA

CB-NMA

- Lu and Ades (2004, 2006, 2009) and many others
- Main Estimand: conditional OR
- Model relative effects (e.g. lnOR)
- Fixed study intercepts
- Assumptions:
  - C-MAR
  - Relative effects are exchangeable
- Cannot include single-arm trials
- Preserve randomization

AB-NMA

- Zhang et al (2014, 2017), Lin et al. (2016, 2017), Hong et al. (2016), Wang et al. (2020, 2021a)
- Main Estimand: conditional or marginal RD/RR/OR
- Model absolute effects with any link functions
- Random study intercepts
- Assumptions:
  - A-MAR
  - Studies are exchangeable
- Naturally include single-arm trials (Wang et al 2021b)
- **May break randomization**

See further comparisons and discussions by Dias & Ades (2016) and White et al. (2019)

Network Meta-analysis (NMA):

Contrast-Based (CB) NMA vs. Arm-Based (AB) NMA

White et al. (2019): "The marginal estimands discussed above use the average underlying risk of the studies in the NMA, which is unlikely to be representative of the target population. External information about clinical populations is therefore valuable for such an analysis. Dias and Ades (2016) argued that, while the overall intervention effect is best estimated in the NMA data set (because randomization promotes internal validity), the overall outcome prevalence is best estimated from clinical registries or other observational sources external to the NMA data set. Any of the models can be used in conjunction with external information to estimate the marginal effect of treatment in a well-defined population."

CB-NMA

- Other estimands are estimated assuming that OR is **transportable/independent** of the baseline prevalence in a population

AB-NMA

Other estimands are estimated assuming **correlation** between treatment-specific event rates with the baseline prevalence in a population

Non-collapsibility of Odds Ratio (OR)

An illustration example to demonstrate the collapsibility of the Risk Difference (RD) and Relative Risk (RR), and non-collapsibility of OR between outcome (Y), treatment (X) and strata (Z).

	Z=1		Z=0		Crude	
	X=1	X=0	X=1	X=0	X=1	X=0
Y=1	80	60	40	20	120	80
Y=0	20	40	60	80	80	120
Risk	0.80	0.60	0.40	0.20	0.60	0.40
RD	0.20		0.20		0.20	
RR	1.33		2.00		1.50	
OR	2.67		2.67		2.25	

- RR varies across the two strata, but it is collapsible across Z as it can be computed as the ratio of weighted average risks.

Non-collapsibility of Odds Ratio (OR)

If a target population is comprised with multiple subpopulations with different baseline risks, then non-collapsibility of OR suggests that other estimands estimated assuming that OR is **transportable or independent** of the subgroup-specific baseline risks vs. the overall baseline risk may differ for the target population. Thus, it may not have a good interpretation.

Whitcomb BW, Naimi AI. Defining, Quantifying, and Interpreting "Noncollapsibility" in Epidemiologic Studies of Measures of "Effect". *Am J Epidemiol* 2021; **190**(5): 697-700.

Q_1: Is the OR Transportable in MA?

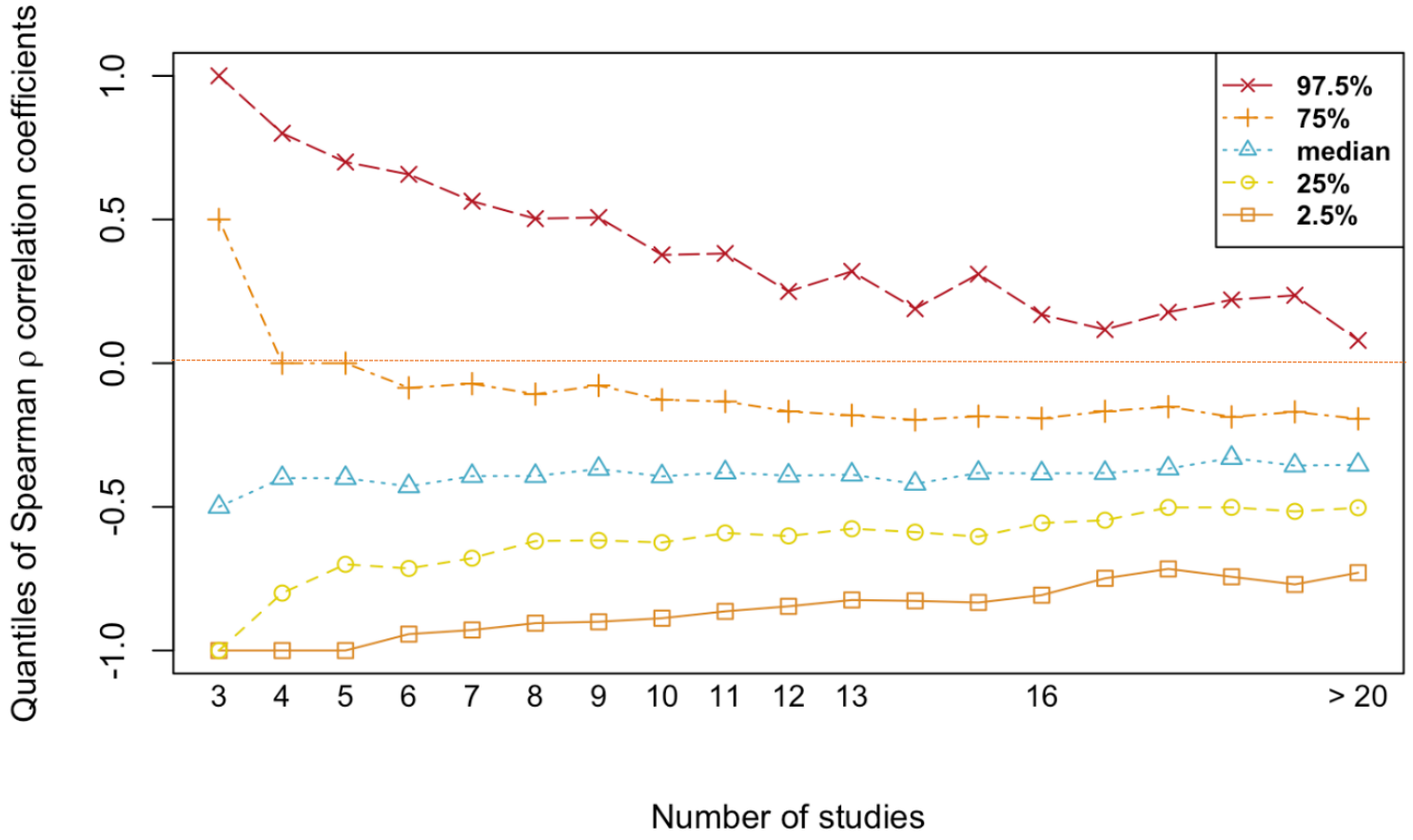


Figure: Quantiles of the Spearman's rank correlation coefficient ρ between the odds ratio and baseline risk among 40,243 meta-analyses, stratified by number of studies in a meta-analysis. (Xiao et al. JCE 2022)

Three Hypothetical Case Studies:

Table *1*. Hypothetical data under constant risk ratio (RR) assumption with 1000 subjects in each arm and fixed baseline risks (in Arm A)

The orange-colored cells are assumed not available in the observed data and assumed available in the full data

Trial	Treatment A	Treatment B	Treatment C
1	60	120	180
2	70	140	210
3	80	160	240
4	90	180	270
5	100	200	300
6	110	220	330
7	120	240	360
8	130	260	390
9	140	280	420
10	150	300	450
11	160	320	480
12	170	340	510
13	180	360	540
14	190	380	570
15	200	400	600

Three Hypothetical Case Studies:

Table *2*. Hypothetical data under constant risk difference (RD) assumption with 1000 subjects in each arm and fixed baseline risks (in Arm A)

The orange-colored cells are assumed not available in the observed data and assumed available in the full data

Trial	Treatment A	Treatment B	Treatment C
1	60	110	160
2	70	120	170
3	80	130	180
4	90	140	190
5	100	150	200
6	110	160	210
7	120	170	220
8	130	180	230
9	140	190	240
10	150	200	250
11	160	210	260
12	170	220	270
13	180	230	280
14	190	240	290
15	200	250	300

Three Hypothetical Case Studies:

Table *3*. Hypothetical data under constant odds ratio (OR) assumption with 1000 subjects in each arm and fixed baseline risks (in Arm A)

The orange-colored cells are assumed not available in the observed data and assumed available in the full data

Trial	Treatment A	Treatment B	Treatment C
1	60	113	161
2	70	131	184
3	80	148	207
4	90	165	229
5	100	182	250
6	110	198	270
7	120	214	290
8	130	230	310
9	140	246	328
10	150	261	346
11	160	276	364
12	170	291	381
13	180	305	397
14	190	319	413
15	200	333	429

Q_2: Can data help us to choose AB vs CB models?

Table *4*. Summary of DIC under constant RR, RD and OR assumptions

		Observed data	Full data
Fixed RR	AB-hom[1]	51	69
	AB-het[2]	**42**	**44**
	CB[3]	50	69
Fixed RD	AB-hom[1]	48	60
	AB-het[2]	**42**	**43**
	CB[3]	47	59
Fixed OR	AB-hom[1]	38	38
	AB-het[2]	42	42
	CB[3]	**36**	**36**

[1]The arm-based NMA with homogeneous variance and equal correlation assumption.
[2]The arm-based NMA with heterogeneous variance and equal correlation assumption.
[3]The contrast-based NMA with homogeneous variance.
[a] Lower deviance information criterion (DIC) indicates better fit

Q_3: Can CB- and AB-NMA Estimate Marginal Effects Accurately Given External Data?

Figure 1. The conditional treatment effects comparing treatment B versus the baseline treatment A for the observed and full data, **based on arm-based NMA with heterogeneous variance and equal correlation assumption**. The three rows correspond to the constant RR, RD and OR assumptions, respectively. The red arrows indicate the magnitude of bias due to model misspecification.

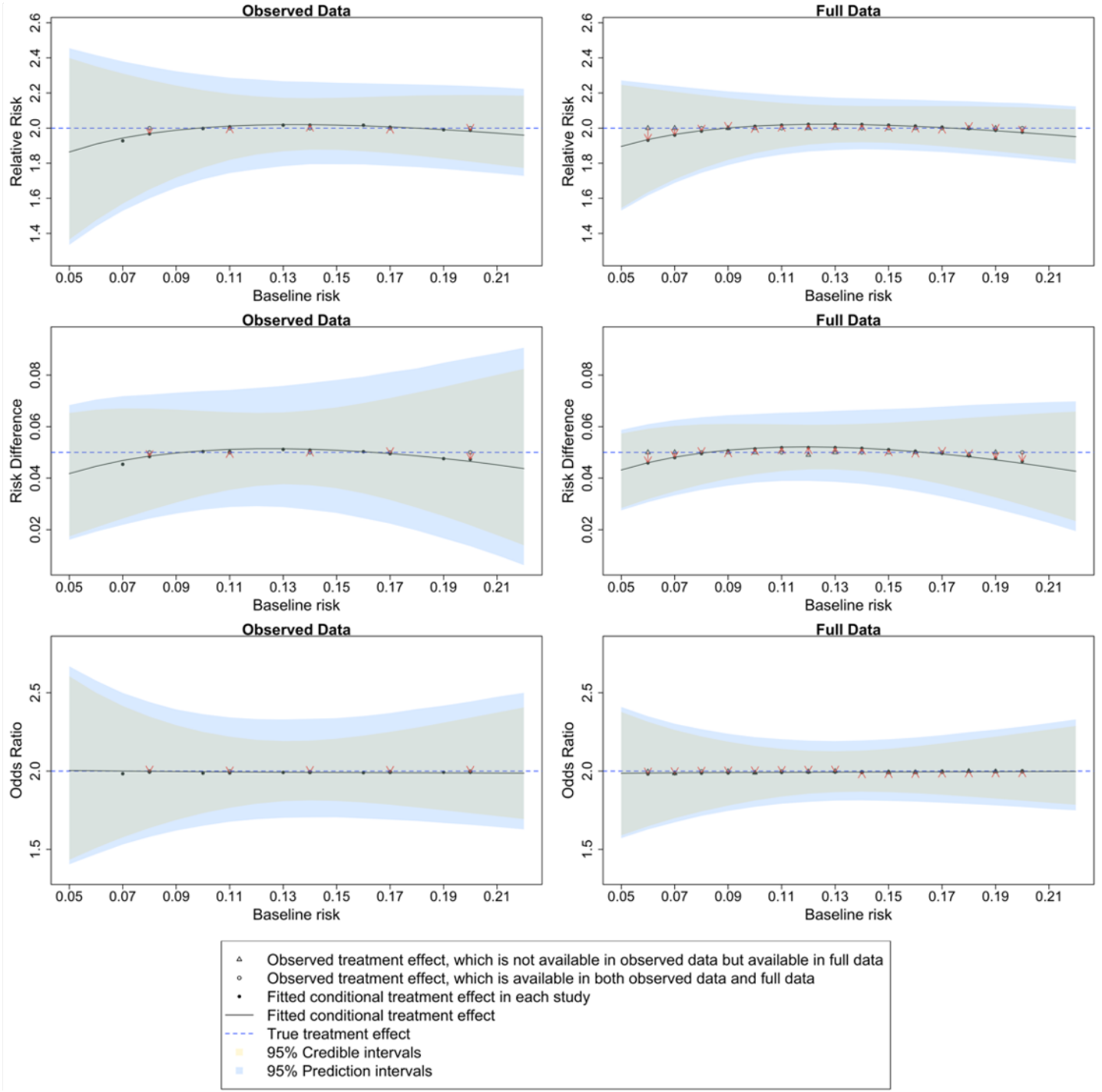
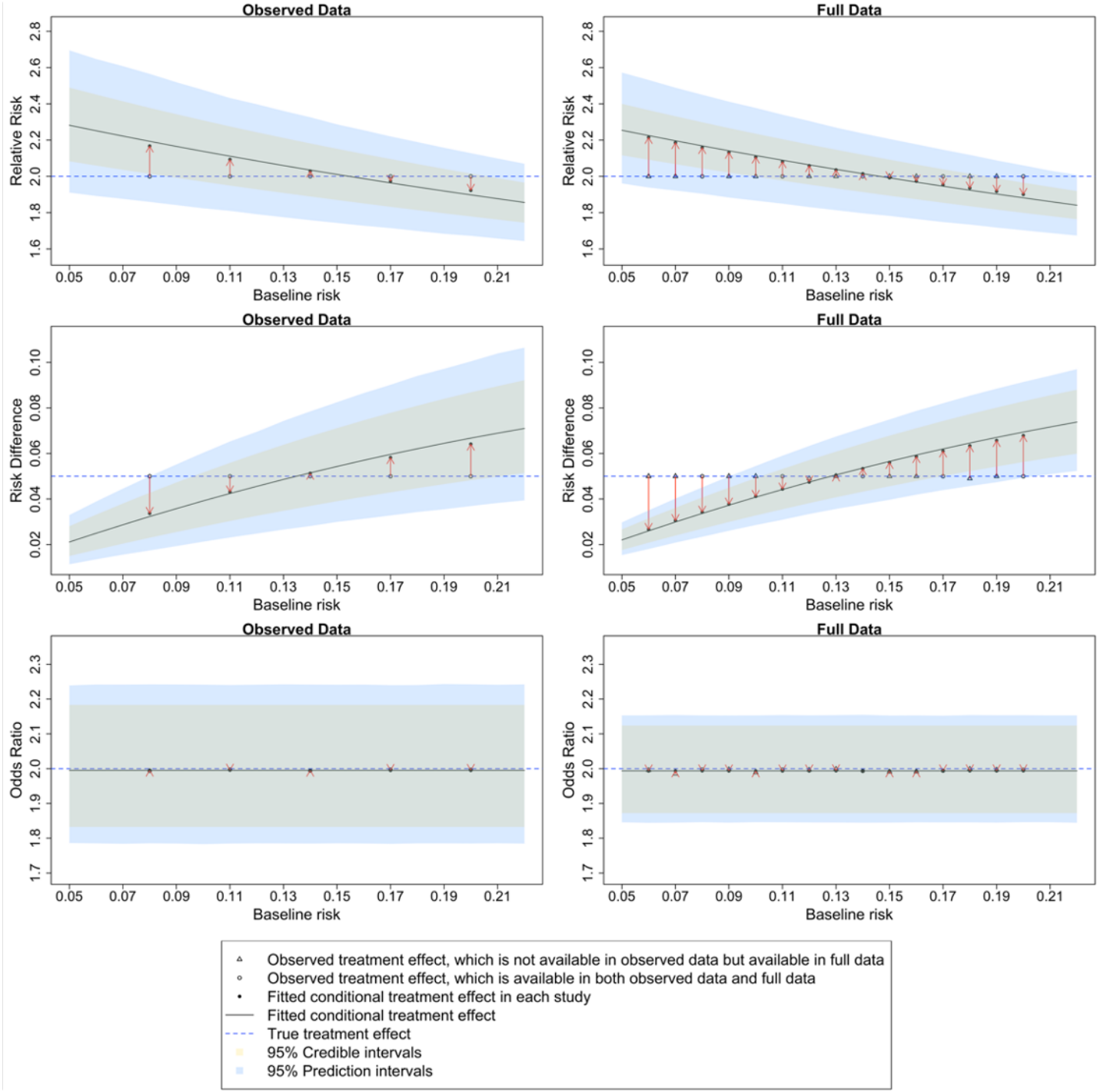


Figure 2. The conditional treatment effects comparing treatment B versus the baseline treatment A for the observed and full data, **based on the contrast-based NMA with homogeneous variance**. The three rows correspond to the constant RR, RD and OR assumptions, respectively. The red arrows indicate the magnitude of bias due to model misspecification.



Take Home Messages: CB-NMA vs. AB-NMA

- Multiple models can give similar goodness-of-fit, suggesting sensitivity analyses using different models. In some cases, data may help us to choose one model over the other, which can lead to better inference.
- As we all know that it is dangerous to extrapolate beyond the scope of the model, it can be dangerous to estimate other estimands assuming that OR is independent of the baseline prevalence in a target population using CB-NMA. At a minimum, one should present the range of baseline prevalence for which the average of conditional OR is estimated.
- AB-NMA is an attractive alternative approach estimating various estimands. If AB-NMA and CB-NMA give different inference, it suggests that some assumptions are not valid, e.g. treatments included in the network might have been examined in different subpopulations and the transitivity assumption may not hold.

Population adjusted indirect comparison (PAIC): MAIC and STC (IPD in Index Trial & AD in Comparator's Trial)

Note: If you have multiple IPDs and some ADs, you may consider **multi-level network meta-regression (ML-NMR)** to synthesize evidence from network of studies (**not discussed**).

J. R. Statist. Soc. A (2020)
183, *Part* 3, *pp.* 1189–1210

Multilevel network meta-regression for population-adjusted treatment comparisons

David M. Phillippo,
University of Bristol, UK

Sofia Dias,
University of York and University of Bristol, UK

A. E. Ades,
University of Bristol, UK

Mark Belger,
Eli Lilly and Company, Windelsham, UK

Alan Brnabic,
Eli Lilly and Company, Sydney, Australia

Alexander Schacht and Daniel Saure,
Eli Lilly and Company, Bad Homburg, Germany

Zbigniew Kadziola
Eli Lilly and Company, Vienna, Austria

and Nicky J. Welton
University of Bristol, UK

Original Research Article

MDM Medical Decision Making

Medical Decision Making
2023, Vol. 43(1) 53–67

DOI: 10.1177/0272989X221117162
journals.sagepub.com/home/mdm
SAGE

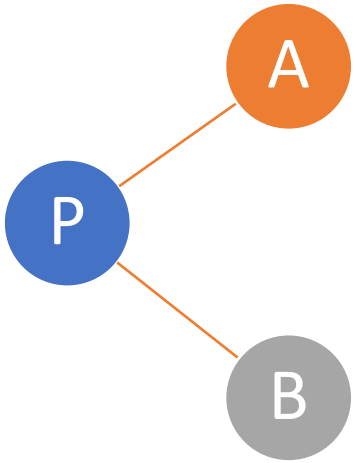
Validating the Assumptions of Population Adjustment: Application of Multilevel Network Meta-regression to a Network of Treatments for Plaque Psoriasis

David M. Phillippo, Sofia Dias, A. E. Ades, Mark Belger, Alan Brnabic, Daniel Saure, Yves Schymura, and Nicky J. Welton

Background. Network meta-analysis (NMA) and indirect comparisons combine aggregate data (AgD) from multiple studies on treatments of interest but may give biased estimates if study populations differ. Population adjustment methods such as multilevel network meta-regression (ML-NMR) aim to reduce bias by adjusting for differences in study populations using individual patient data (IPD) from 1 or more studies under the conditional constancy assumption. A shared effect modifier assumption may also be necessary for identifiability. This article aims to demonstrate how the assumptions made by ML-NMR can be assessed in practice to obtain reliable treatment effect estimates in a target population. **Methods.** We apply ML-NMR to a network of evidence on treatments for plaque psoriasis with a mix of IPD and AgD trials reporting ordered categorical outcomes. Relative treatment effects are estimated for each trial population and for 3 external target populations represented by a registry and 2 cohort studies. We examine residual heterogeneity and inconsistency and relax the shared effect modifier assumption for each covariate in turn. **Results.** Estimated population-average treatment effects were similar across study populations, as differences in the distributions of effect modifiers were small. Better fit was achieved with ML-NMR than with NMA, and uncertainty was reduced by explaining within- and between-study variation. We found little evidence that the conditional constancy or shared effect modifier assumptions were invalid. **Conclusions.** ML-NMR extends the NMA framework and addresses issues with previous population adjustment approaches. It coherently synthesizes evidence from IPD and AgD studies in networks of any size while avoiding aggregation bias and noncollapsibility bias, allows for key assumptions to be assessed or relaxed, and can produce estimates relevant to a target population for decision-making.

Motivation for population adjusted indirect comparisons

- Let consider a simple star-shaped network: 1) in the AP trial (labelled AvP), we can estimate the relative effect $\hat{d}_{AP(AvP)}$; 2) in the BP trial (labelled BvP), we can estimate the relative effect $\hat{d}_{BP(BvP)}$. The Bucher's method and standard network meta-analysis makes the consistency assumption so the effect of A versus B would be $\hat{d}_{AB(T)} = \hat{d}_{AP(AvP)} - \hat{d}_{BP(BvP)}$.



- However, $\hat{d}_{AB(T)}$ is not specific to a target population (labelled T), and biased if the transitivity assumption does not hold. Intuitively, $\hat{d}_{AB(T)}$ would be valid for some mixture of the *AvP* and *BvP* population.
- Population adjusted methods, unlike NMA, aim to estimate a relative effect for a specific population: (AvP), (BvP), or other target population.

PAIC assumptions to control for population differences:

- Studies must have similar designs, e. g. same outcome definitions. PAIC methods cannot adjust for structural difference between studies.
- The index trial population should be broader than the comparator trial populations on all important baseline characteristics (including prognostic factors and effect modifiers).
- For example, if index trial had excluded vaccinated subjects, we would have no information on treatment outcomes in vaccinated subjects, thus not possible to adjust for differences in vaccine status in a comparison against a comparator trial if it contained vaccinated subjects. In this situation, we would require:
  - ✓ the fraction of vaccinated subjects in a comparator trial is small enough and therefore has a negligible effect on outcomes, or
  - ✓ Vaccine only has a negligible effect on outcomes.

Anchored PAIC versus Unanchored PAIC:

Both Matching Adjusted Indirect Comparison (MAIC) *and* Simulated Treatment Comparison (STC) *can be used to carry out either an "anchored" indirect comparison, where there is a common comparator arm in each trial, or an "unanchored" indirect comparison, where there is a disconnected treatment network or single-arm studies. An unanchored MAIC or STC effectively assumes that absolute outcomes can be predicted from the covariates; that is, it assumes that all effect modifiers and prognostic factors are accounted for. This assumption is very strong, and largely considered impossible to meet.*

NICE DSU Technical Support Document (#18) .

Anchored PAIC versus Unanchored PAIC:

The NICE recommendations:

- Anchored comparison may be considered when there is connected evidence with a common comparator. Unanchored comparisons may only be considered where single-arm studies are involved, or in the absence of a connected network of randomized evidence.
- Submissions using **anchored** population adjusted analyses need to provide evidence that they are less likely to produce biased estimates of treatment difference than could be achieved through standard methods, and that population adjustment would have a material impact on relative effect estimates due to the removal of substantial bias.
- Submissions using **unanchored** population adjusted analyses need to provide evidence that absolute outcomes can be predicted with sufficient accuracy in relation to the relative treatment effects.

NICE DSU Technical Support Document (#18) .

Anchored PAIC versus Unanchored PAIC:

The NICE recommendations:

- Outcome regression methods should adjust for all effect modifiers and any prognostic variables that improve model fit in **anchored** comparisons. In **unanchored** comparisons, all effect modifiers and prognostic factors should be adjusted for, in order to reliably predict absolute outcomes. In practice, these requirement rarely hold.
- Indirect comparisons should be carried out on the linear predictor scale.

NICE DSU Technical Support Document (#18) .

Current Debate on MAIC vs. STC

Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study

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

[1]Bristol Medical School (Population Health Sciences), University of Bristol, Bristol, UK

[2]Centre for Reviews and Dissemination, University of York, York, UK

Correspondence
David M. Phillippo, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK.
Email: david.phillippo@bristol.ac.uk

Funding information
UK Medical Research Council, Grant/Award Numbers: MR/P015298/1, MR/R025223/1

Standard network meta-analysis and indirect comparisons combine aggregate data from multiple studies on treatments of interest, assuming that any factors that interact with treatment effects (effect modifiers) are balanced across populations. Population adjustment methods such as multilevel network meta-regression (ML-NMR), matching-adjusted indirect comparison (MAIC), and simulated treatment comparison (STC) relax this assumption using individual patient data from one or more studies, and are becoming increasingly prevalent in health technology appraisals and the applied literature. Motivated by an applied example and two recent reviews of applications, we undertook an extensive simulation study to assess the performance of these methods in a range of scenarios under various failures of assumptions. We investigated the impact of varying sample size, missing effect modifiers, strength of effect modification and validity of the shared effect modifier assumption, validity of extrapolation and varying between-study overlap, and different covariate distributions and correlations. ML-NMR and STC performed similarly, eliminating bias when the requisite assumptions were met. Serious concerns are raised for MAIC, which performed poorly in nearly all simulation scenarios and may even increase bias compared with standard indirect comparisons. All methods incur bias when an effect modifier is missing, highlighting the necessity of careful selection of potential effect modifiers prior to analysis. When all effect modifiers are included, ML-NMR and STC are robust techniques for population adjustment. ML-NMR offers additional advantages over MAIC and STC, including extending to larger treatment networks and producing estimates in any target population, making this an attractive choice in a variety of scenarios.

KEYWORDS
effect modification, indirect comparison, individual patient data, matching-adjusted indirect comparison, multilevel network meta-regression, simulated treatment comparison

Methods for population adjustment with limited access to individual patient data: A review and simulation study

Antonio Remiro-Azócar[1,2] | Anna Heath[1,3,4] | Gianluca Baio[1]

[1]Department of Statistical Science, University College London, London, UK

[2]Quantitative Research, Statistical Outcomes Research & Analytics (SORA) Ltd., London, UK

[3]Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada

[4]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

Correspondence
Antonio Remiro-Azócar, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK.
Email: antonio.remiro.16@ucl.ac.uk

Funding information
Canadian Institutes of Health Research, Grant/Award Number: MYG-151207; Engineering and Physical Sciences Research Council; Mapi/ICON

Abstract
Population-adjusted indirect comparisons estimate treatment effects when access to individual patient data is limited and there are cross-trial differences in effect modifiers. Popular methods include matching-adjusted indirect comparison (MAIC) and simulated treatment comparison (STC). There is limited formal evaluation of these methods and whether they can be used to accurately compare treatments. Thus, we undertake a comprehensive simulation study to compare standard unadjusted indirect comparisons, MAIC and STC across 162 scenarios. This simulation study assumes that the trials are investigating survival outcomes and measure continuous covariates, with the log hazard ratio as the measure of effect. MAIC yields unbiased treatment effect estimates under no failures of assumptions. The typical usage of STC produces bias because it targets a conditional treatment effect where the target estimand should be a marginal treatment effect. The incompatibility of estimates in the indirect comparison leads to bias as the measure of effect is non-collapsible. Standard indirect comparisons are systematically biased, particularly under stronger covariate imbalance and interaction effects. Standard errors and coverage rates are often valid in MAIC but the robust sandwich variance estimator underestimates variability where effective sample sizes are small. Interval estimates for the standard indirect comparison are too narrow and STC suffers from bias-induced undercoverage. MAIC provides the most accurate estimates and, with lower degrees of covariate overlap, its bias reduction outweighs the loss in precision under no failures of assumptions. An important future objective is the development of an alternative formulation to STC that targets a marginal treatment effect.

KEYWORDS
clinical trials, comparative effectiveness research, health technology assessment, indirect treatment comparison, oncology, simulation study

Current Debate on MAIC vs. STC

Received: 3 November 2020 | Accepted: 1 December 2020
DOI: 10.1002/sim.8857

LETTER TO THE EDITOR

Statistics in Medicine WILEY

Conflating marginal and conditional treatment effects: Comments on "Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study"

Antonio Remiro-Azócar[1] | Anna Heath[1,2,3] | Gianluca Baio[1]

[1]Department of Statistical Science, University College London, London, UK
[2]Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Canada
[3]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

Correspondence
Antonio Remiro Azócar, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK.
Email: antonio.remiro.16@ucl.ac.uk

Funding information
Canadian Institutes of Health Research, Grant/Award Number: Anna Heath: MYG-151207 Innovative Clinical Trials; Engineering and Physical Sciences Research Council, Grant/Award Number: Antonio Remiro-Azocar: PhD scholarship

In this commentary, we highlight the importance of: (1) carefully considering and clarifying whether a marginal or conditional treatment effect is of interest in a population-adjusted indirect treatment comparison; and (2) developing distinct methodologies for estimating the different measures of effect. The appropriateness of each methodology depends on the preferred target of inference.

KEYWORDS
causal inference, conditional treatment effect, indirect treatment comparison, marginal treatment effect, population adjustment

Received: 1 December 2020 | Accepted: 1 December 2020
DOI: 10.1002/sim.8965

AUTHORS REPLY

Statistics in Medicine WILEY

Target estimands for efficient decision making: Response to comments on "Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study"

David M. Phillippo[1] | Sofia Dias[1,2] | Anthony E. Ades[1] | Nicky J. Welton[1]

[1]Bristol Medical School (Population Health Sciences), University of Bristol, Bristol, UK
[2]Centre for Reviews and Dissemination, University of York, York, UK

Correspondence
David M. Phillippo, Bristol Medical School (Population Health Sciences), University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK.
Email: david.phillippo@bristol.ac.uk

Funding information
Medical Research Council, Grant/Award Numbers: MR/P015298/1, MR/R025223/1

We thank Remiro-Azócar, Heath, and Baio (R-AHB) for their letter to the editor,[1] in response to our recent article presenting a simulation study comparing the performance of methods for population-adjusted indirect comparison.[2] R-AHB discuss the important issue of target estimands with noncollapsible effect measures, expanding upon the discussion in sections 4.3 and 7 of our article.[2] R-AHB distinguish between marginal and conditional treatment effect estimates and explain that matching-adjusted indirect comparison (MAIC) targets marginal effects whereas simulated treatment comparison (STC) and multilevel network meta-regression (ML-NMR) target conditional treatment effects. They conclude that "methods like MAIC are valid for population-based inference, but not "fit for purpose" when inference is at the individual level, whereas methods like ML-NMR are valid for inference at the individual level, but not designed for population-based inference." Furthermore, they assert that marginal treatment effect estimates are necessary for population-based inference as required for decision-making in Health Technology Assessment (HTA).

We welcome and encourage debate of these issues, which—despite much discussion in the literature on randomized controlled trials (RCTs)[3-6] and observational epidemiology[7-9]—have largely been overlooked in the literature on population adjustment and meta-analysis to date. However, whilst we agree with R-AHB that population-based inference is required for HTA, we disagree that methods like ML-NMR are not appropriate to obtain population-average estimates for HTA. In this response, we further clarify the use of conditional estimates to inform population-average treatment effects and why we believe these are appropriate target estimands for decision making. We also correct some important inaccuracies in R-AHB's letter regarding the characterization of the methods (in particular ML-NMR) and interpretation of our simulation study results.

Current Debate on MAIC vs. STC

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data

Antonio Remiro-Azócar[1,2] | Anna Heath[1,3,4] | Gianluca Baio[1]

[1]Department of Statistical Science, University College London, London, UK
[2]Quantitative Research, Statistical Outcomes Research & Analytics (SORA) Ltd, London, UK
[3]Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Canada
[4]Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

Correspondence
Antonio Remiro-Azócar, Department of Statistical Science, University College London, London, UK.
Email: antonio.remiro.16@ucl.ac.uk

Funding information
Canada Research Chairs, Grant/Award Number: Anna Heath: Statistical Trial Design; Engineering and Physical Sciences Research Council, Grant/Award Number: Antonio Remiro-Azocar: PhD scholarship; Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: Anna Heath: Discovery Grant Program RGPIN-2021-03

Abstract

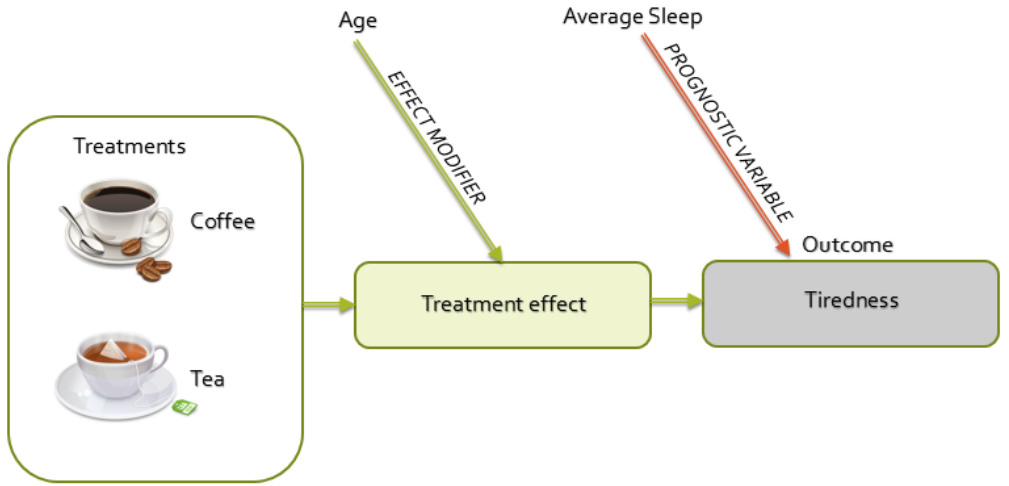
Population adjustment methods such as matching-adjusted indirect comparison (MAIC) are increasingly used to compare marginal treatment effects when there are cross-trial differences in effect modifiers and limited patient-level data. MAIC is based on propensity score weighting, which is sensitive to poor covariate overlap and cannot extrapolate beyond the observed covariate space. Current outcome regression-based alternatives can extrapolate but target a conditional treatment effect that is incompatible in the indirect comparison. When adjusting for covariates, one must integrate or average the conditional estimate over the relevant population to recover a compatible marginal treatment effect. We propose a marginalization method based on parametric G-computation that can be easily applied where the outcome regression is a generalized linear model or a Cox model. The approach views the covariate adjustment regression as a nuisance model and separates its estimation from the evaluation of the marginal treatment effect of interest. The method can accommodate a Bayesian statistical framework, which naturally integrates the analysis into a probabilistic framework. A simulation study provides proof-of-principle and benchmarks the method's performance against MAIC and the conventional outcome regression. Parametric G-computation achieves more precise and more accurate estimates than MAIC, particularly when covariate overlap is poor, and yields unbiased marginal treatment effect estimates under no failures of assumptions. Furthermore, the marginalized regression-adjusted estimates provide greater precision and accuracy than the conditional estimates produced by the conventional outcome regression, which are systematically biased because the measure of effect is non-collapsible.

KEYWORDS
causal inference, health technology assessment, indirect treatment comparison, marginal treatment effect, outcome regression, standardization

Matching Adjusted Indirect Comparison (MAIC)

Adjust the differences between populations by weighting patients in the index trial so that the average characteristics (and SDs if continuous) match the comparator's population.

- Compare the characteristics of the IPD trial to be weighted against the target population
  - *Effect modifiers refer to characteristics that impact the relative treatment effect*
  - *Prognostic variables refer to characteristics that directly affect the outcome, but not the relative treatment effect on the outcome.*



- Calculate and check the MAIC weights
  - NICE provides detailed instruction and R code – http://nicedsu.org.uk/technical-support-documents/population-adjusted-indirect-comparisons-maic-and-stc/
  - R MAIC package https://cran.r-project.org/web/packages/maic/index.html.
- Check the balance of the weighted trials
- Compare the weighted and the unweighted outcome, and evaluate the uncertainty

Matching Adjusted Indirect Comparison (MAIC)

To estimate patient weights through PS, MAIC must match both **effect modifiers and prognostic variables** in the index trial to that of the comparator's population: **Method of Moments Weights**

- **The objective** is to ensure that the means [and SDs if continuous] of the covariates of the reweighted index trial patients match the means [and SDs] of the comparator's population.
- IPD in the index trial: covariate vector $\boldsymbol{X_i}$, with $i = 1, \dots, n$, where $\boldsymbol{n}$ **is # of patients**.
- AD in the comparator's population: the means [and standard deviations (SDs) if continuous variables, as $\mathrm{Var(X)} = E(X^2) + E^2(X)$, create additional covariates X^2] of the covariates as vector $\boldsymbol{\overline{X}_P}$.
- The weight, which is equivalent to a propensity score, assigned to patient i is $\widehat{\omega}_i = \frac{\exp(\boldsymbol{X_i^T}\widehat{\boldsymbol{\theta}})}{\sum_{i=1}^{n}\exp(\boldsymbol{X_i^T}\widehat{\boldsymbol{\theta}})}$ with the vector $\widehat{\boldsymbol{\theta}}$ estimated as the solution to:

$$\frac{\sum_{i=1}^{n}\boldsymbol{X}_i\exp(\boldsymbol{X_i^T}\widehat{\boldsymbol{\theta}})}{\sum_{i=1}^{n}\exp(\boldsymbol{X_i^T}\widehat{\boldsymbol{\theta}})} - \boldsymbol{\overline{X}_P} = 0 \Leftrightarrow \frac{\sum_{i=1}^{n}\boldsymbol{Z}_i\exp(\boldsymbol{Z_i^T}\widehat{\boldsymbol{\theta}})}{\sum_{i=1}^{n}\exp(\boldsymbol{Z_i^T}\widehat{\boldsymbol{\theta}})} = 0, \text{ where } \boldsymbol{Z}_i = \boldsymbol{X}_i - \boldsymbol{\overline{X}_P}$$

- This estimator has unique solutions and converges to the true value for $\boldsymbol{\theta}$.

Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, et al. Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. Value in Health. 2012;15:940-7.
Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, et al. Comparative Effectiveness Without Head-to-Head Trials. Pharmacoeconomics, Methodological Considerations. 2010;28(10):935-45.

Matching Adjusted Indirect Comparison (MAIC):

Estimating treatment effects on MAIC reweighted populations

- To estimate treatment effects on MAIC reweighted populations, we denote the outcome for individual i in index trial as Y_i and use the estimated weights $\hat{\omega}_i$ to form the weighted estimator in the comparator's population P

$$\hat{Y}_{(P)} = \frac{\sum_{i=1}^{n} \hat{\omega}_i Y_i}{\sum_{i=1}^{n} \hat{\omega}_i}$$

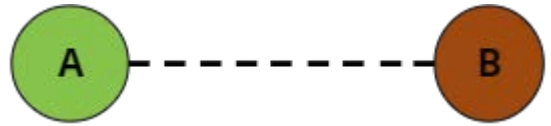
- Although it is a simple weighted mean, we use a weighted (generalized) linear model to correctly calculate standard errors using **robust sandwich estimators**.
- Sandwich estimators are derived empirically from the data rather than making overly strong assumptions about the weights, to account for the fact that the weights $\hat{\omega}_i$ are estimated rather than fixed and known.
- Effective sample size ($\boldsymbol{ESS}$) $= (\sum_{i=1}^{n} \hat{\omega}_i)^2 / \sum_{i=1}^{n} \hat{\omega}_i^{\,2}$.

Matching Adjusted Indirect Comparison (MAIC): Connection to Calibration Estimation

- In sample survey, the calibration estimation choose weights that match the means (and standard deviations if continuous), i.e., $\sum_{i=1}^{n} w_i \boldsymbol{X_i} = \overline{\boldsymbol{X}}_{\boldsymbol{P}}$, by minimizing some objective function $\sum_{i=1}^{n} D(w_i)$ with $\sum_{i=1}^{n} w_i = 1$. In general, $D(w_i)$ is a distance between w_i and the uniform weights $1/n$.
- The entropy balance weights use the entropy distance, $D(w_i) = w_i \log w_i$, which is equivalent to method of moments weights.
- Other types of distance such as the quadratic distance or the absolute distance can be applied as well.
- **MAIC with maximum ESS** is equivalent to the calibration estimation minimizing quadratic distance $\sum_{i=1}^{n} D(w_i) = \sum_{i=1}^{n} (w_i - 1/n)^2 = \sum_{i=1}^{n} w_i^2 - \frac{2}{n}\sum_{i=1}^{n} w_i + \frac{1}{n^2}$.

Jackson D et al. Alternative weighting schemes when performing matching-adjusted indirect comparisons.

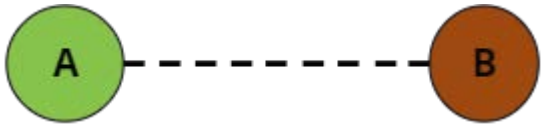
Simulated Treatment Comparison (STC)



STC involves estimating an outcome regression model for the relationship between population characteristics and outcome in an index trial with IPD, and then using the model to estimate that outcome for the comparator's trial population (with only AD).

- Fit an outcome regression model in the index trial $\mathrm{g}\left(\mu_{\boldsymbol{X_{iA}}}|\mathrm{A}\right) = \boldsymbol{X_{iA}\beta};$
- *Estimate the outcome of treatment A on the population of comparator's trial B by substituting in the mean covariate values of trial B as* $\mathrm{g}\left(\mu_{\bar{\boldsymbol{X}}_{\boldsymbol{B}}}|B\right) = \bar{\boldsymbol{X}}_{\boldsymbol{B}}\boldsymbol{\beta};$
- Indirect comparisons should be carried out on the linear predictor scale, i.e. $\mathrm{g}(\hat{\mu}|B)$ vs $g\left(\mu_{\bar{X}_B}|B\right)$.

Simulated Treatment Comparison (STC)



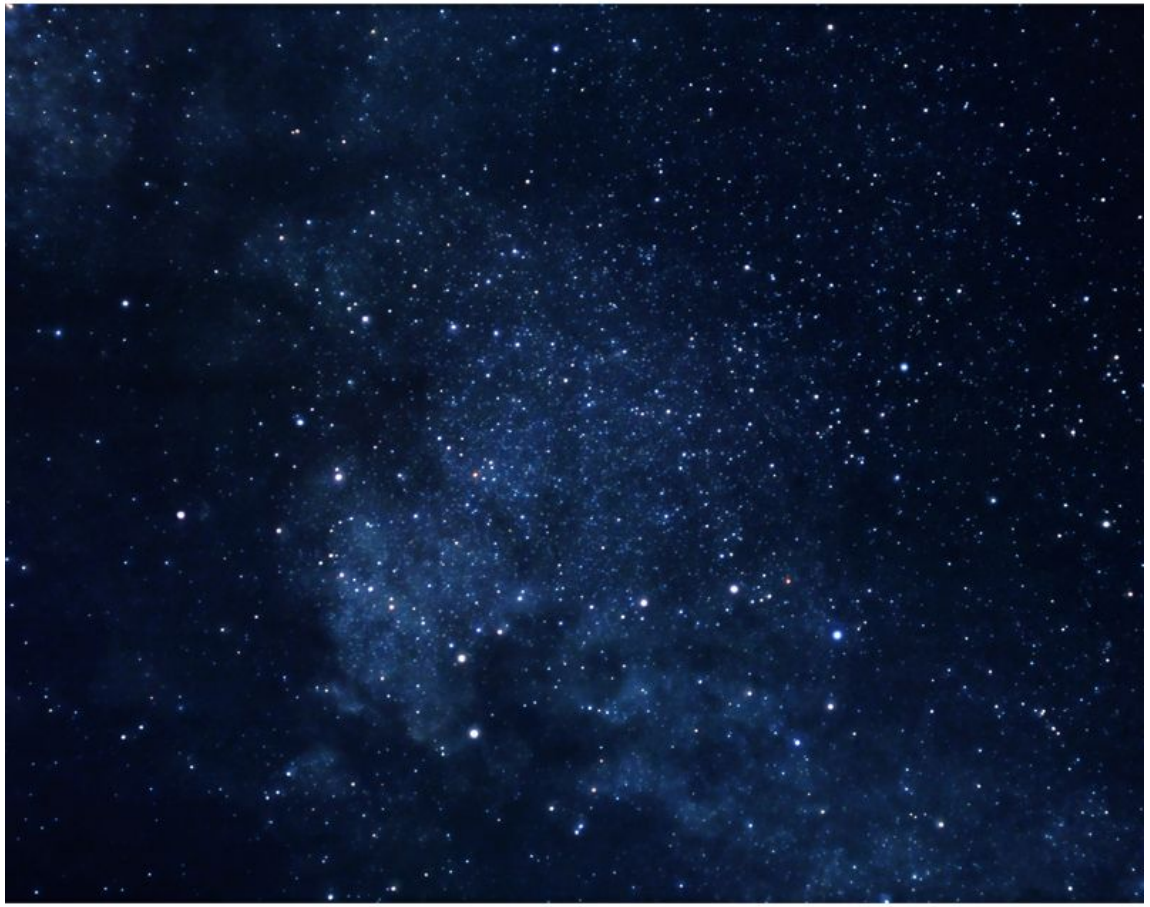
Alternatively, to overcome the issue due to $E(g(X)) \neq g\big(E(X)\big)$ for nonlinear link functions, patient profiles can be simulated to reflect the comparator population by

- Assuming covariance between effect modifiers and prognostic variables in index trial A applies to comparator's trial B.
- Setting means to those from the comparator's trial B and and simulate $\boldsymbol{X_{iB}}$ and predict $\mathrm{g}\big(\mu_{\boldsymbol{X_{iB}}}|B\big) = \boldsymbol{X_{iB}\beta}$ based on the outcome model from trial A
- Carry out the comparison on the marginal outcome $E(\mu_{\boldsymbol{X_{iB}}})$ vs. $\hat{\mu}|B$.

Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data

Conclusion: Indirect Treatment Comparisons: When Are They Needed and How Do They Work?

- NMA is the gold standard for indirect comparisons of multiple treatments, however incomplete evidence networks and heterogeneity (among other things) between studies may limit the use.
- STC and MAIC can overcome these challenges by carrying out a targeted comparison between outcomes for specific treatment arms of interest.
- Statistical adjustment is required to reduce confounding in the comparisons. STCs achieve this with the use of outcome predictive models, while MAIC relies on reweighting subjects.
- In practice, it is important to consider multiple approaches as sensitivity analyses and to provide totality of evidence as indirect treatment comparisons is at high risk of bias no matter which approach is chosen.



Night Sky Constellations

The Big Dipper

Be aware of the limitations of any source of information and potential biases;

Seek out diverse perspectives to avoid misleading conclusions.


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
bedankt
dziękuję
go raibh maith agat
sukriya
kop khun krap
terima kasih
grazie
arigatô
dakujem
спасибо
рахмат
Баярлалаа
mersi
barka
welalin
tack
spas
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
хвала
asante
manana
obrigada
tenki
murakoze
chokrane
mamnun
дякую
djiere dieuf
tau
mochchakkeram
faafetai lava
vinaka
спасибі
blagodaram
kiitos
dankie
dhanyavad
nanni
nandri
bayarlalaa
gràçie
hvala
maururu
köszönöm
enkosi
akun
dankon
aciû
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
sobodi
dëkuji
sagolun
mèsi
didi madloba
najis tuke
ありがとう
tanemirt
rahmet
takk
trugarez
dhanyavadagalu
diolch
shukriya
merci
мерси
ευχαριστώ
kam sah hamnida
rahmat
তোমাকে ধন্যবাদ
감사합니다
xièxie