# Apples and oranges in the context of anchored indirect treatment comparisons – Is there more to it than effect modifiers?

2023 ISPOR Boston, May 10, 10:00 – 11:15 AM ET

PRECISIONheor

UBC

BAYER

UCSF
University of California
San Francisco

# Population adjustment in the context of indirect comparisons

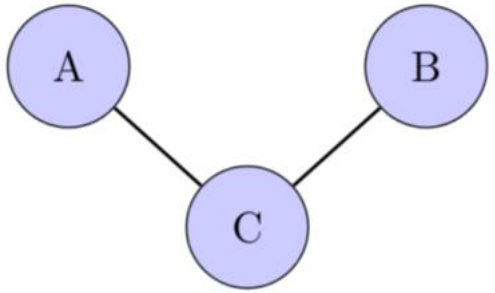
Antonio Remiro-Azócar, PhD
Medical Affairs Statistics, Bayer, UK

# Background

The following setting is common in HTA submissions

- An active treatment (treatment *A*) needs to be compared against a competitor (treatment *B*)
- No head-to-head randomized trial between treatments *A* and *B*
- We have individual patient data (IPD) for study *A* but not for study *B*
- There are differences in baseline characteristics between study *A* and study *B*
- We standardize/marginalize study *A* over the covariate distribution of study *B* for a compatible ITC

ANCHORED COMPARISON




$$\hat{\Delta}_{AB}^{(B)} = \hat{\Delta}_{AC}^{(B)} - \hat{\Delta}_{BC}^{(B)}$$

UNANCHORED COMPARISON


$$\hat{\Delta}_{AB}^{(B)} = g(\hat{\mu}_{A}^{(B)}) - g(\hat{\mu}_{B}^{(B)})$$

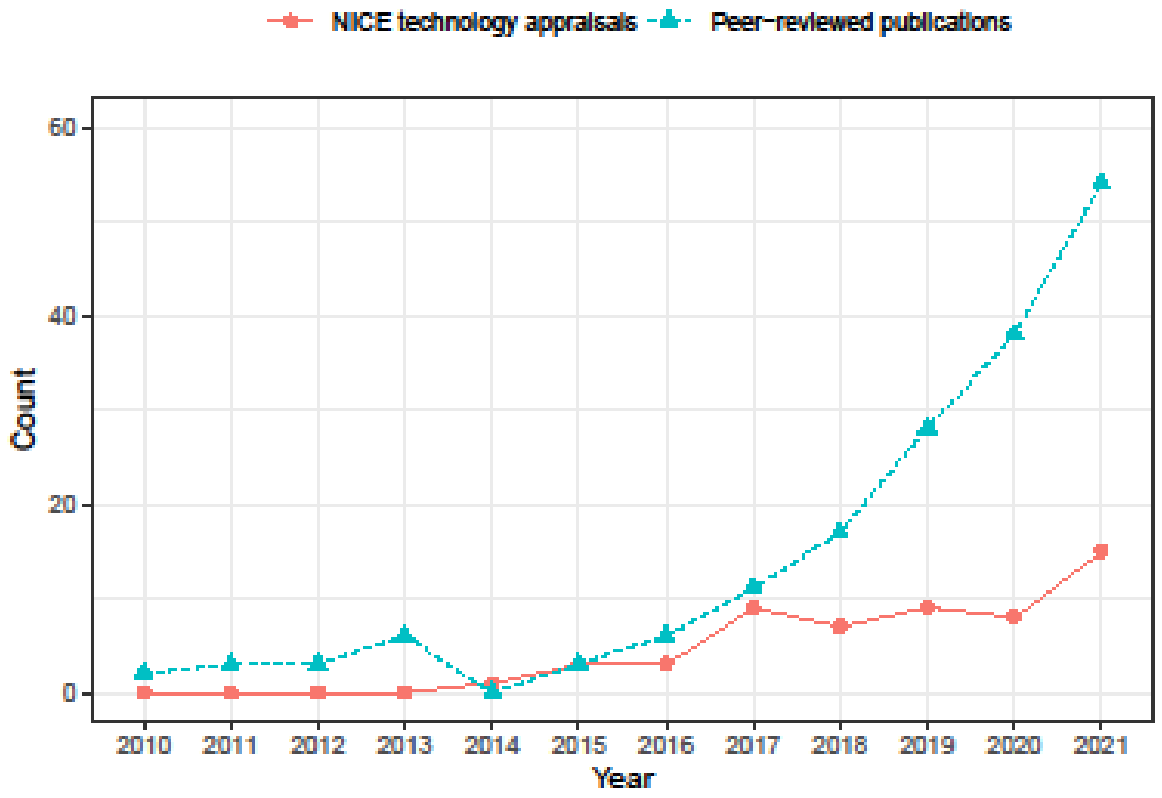
# Covariate-adjusted ITCs (2010-2021)

## Matching-adjusted indirect comparison (MAIC)

- Odds-weighting approach
- 164 peer-reviewed applications
- 50 NICE technology appraisals

## Simulated treatment comparison (STC)

- Outcome modelling approach
- 20 peer-reviewed applications
- 9 NICE technology appraisals



# Matching-adjusted indirect comparison (MAIC)

*Adjusting for between-trial differences by weighting*

- Logistic model for trial assignment

$$\ln(w_i) = \ln\left[\frac{Pr(S = B \mid \boldsymbol{x}_i)}{1 - Pr(S = B \mid \boldsymbol{x}_i)}\right] = \alpha_0 + \boldsymbol{x}_i\boldsymbol{\alpha_1}$$

- Entropy balancing approach; covariate balance is viewed as a convex optimization problem

$$Q(\boldsymbol{\alpha_1}) = \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_i\boldsymbol{\alpha_1}\right)$$

- The estimated weights denote the conditional odds of assignment to study *B*

$$\hat{w}_i = \exp(\boldsymbol{x}_i\hat{\boldsymbol{\alpha}}_{\mathbf{1}})$$

- Marginal mean outcomes and/or relative effects for study *A* treatment(s) estimated in study *B*

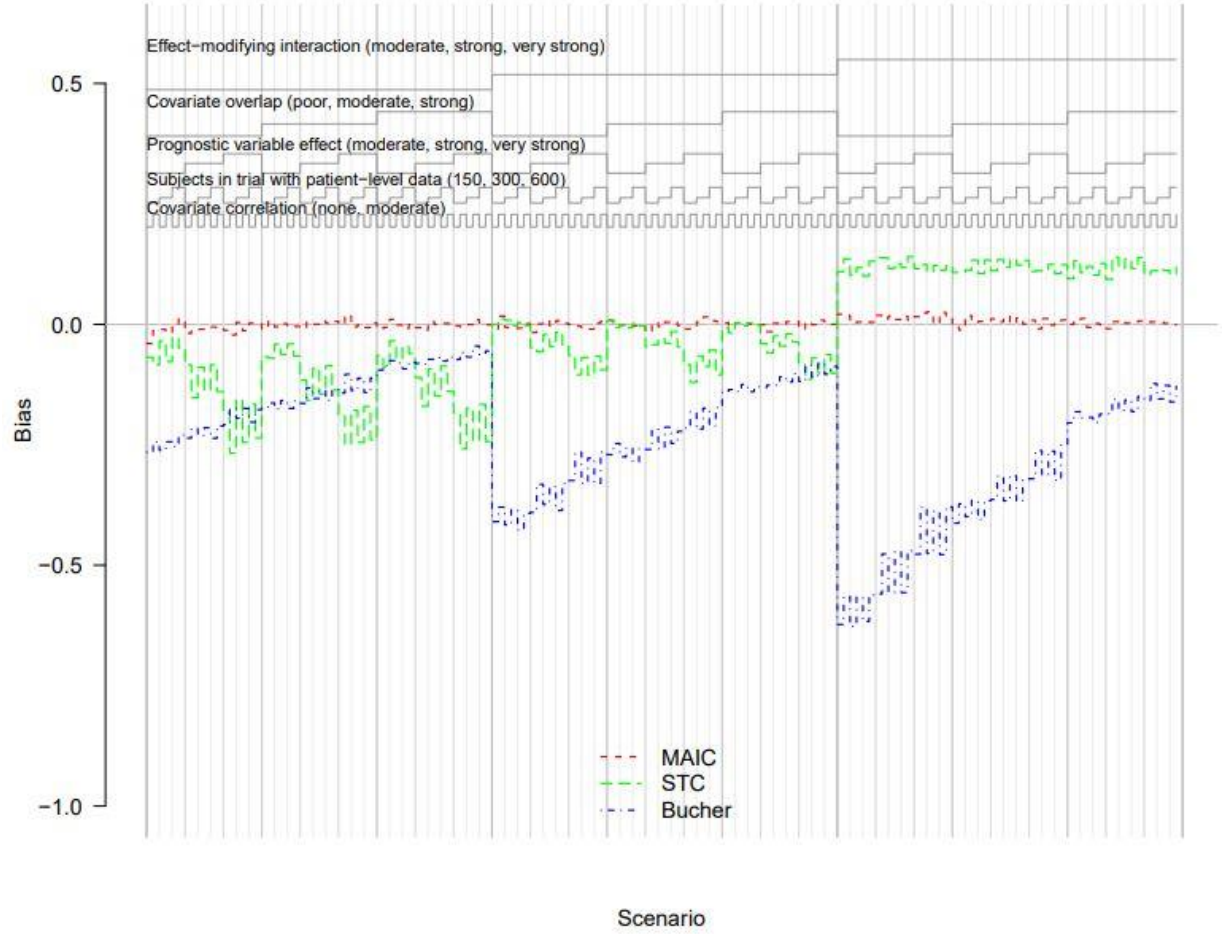
$$\hat{\mu}_t^{(B)} = \frac{\sum_{i=1}^{n_t} y_{i,t}\hat{w}_{i,t}}{\sum_{i=1}^{n_t} \hat{w}_{i,t}}$$

$$\hat{\Delta}_{AC}^{(B)} = g\left(\hat{\mu}_A^{(B)}\right) - g\left(\hat{\mu}_C^{(B)}\right)$$

# Simulated treatment comparison (STC)

*Adjusting for between-trial differences by outcome modelling*

- The version from NICE DSU TSD 18 targets a conditional as opposed to a marginal effect
- This leads to bias for non-collapsible effect measures, e.g. (log) hazard and odds ratios




Remiro-Azócar, A., Heath, A. and Baio, G., 2021. Methods for population adjustment with limited access to individual patient data: A review and simulation study. Research synthesis methods, 12(6), pp.750-775.

# Simulated treatment comparison (STC)

*Parametric model-based standardization or G-computation*

- Simulate individual-level covariates for study *B*, e.g., using a copula distribution

- Fit a multivariable regression of outcome on covariates (and treatment) to the IPD for study *A*

$$g(\mu_i) = \beta_0 + \boldsymbol{x}_i\boldsymbol{\beta_1} + \left(\beta_t + \boldsymbol{x}_i^{(\boldsymbol{EM})}\boldsymbol{\beta_2}\right)\mathbb{1}(t_i = A)$$ (anchored case)

- Use the coefficients of the fitted model to predict hypothetical outcomes under the study *A* treatments for each simulated subject

- Consider the anchored case. We plug each treatment value into the regression fit to compute the marginal outcome means under *A* and *C*, and the corresponding relative effect.

$$\hat{\mu}_A^{(B)} = \frac{1}{m}\sum_{j=1}^{m} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_j^*\hat{\boldsymbol{\beta}}_{\mathbf{1}} + \hat{\beta}_t + \boldsymbol{x}_j^{*(\boldsymbol{EM})}\hat{\boldsymbol{\beta}}_{\mathbf{2}}) \qquad \hat{\mu}_C^{(B)} = \frac{1}{m}\sum_{j=1}^{m} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_j^*\hat{\boldsymbol{\beta}}_{\mathbf{1}})$$

$$\hat{\Delta}_{AC}^{(B)} = g\left(\hat{\mu}_A^{(B)}\right) - g\left(\hat{\mu}_C^{(B)}\right)$$

- A Bayesian implementation is also feasible; good for probabilistic sensitivity analysis

# Weighting or outcome modelling?

*NICE DSU recommendations (anchored scenario)*

CHTE2020 SOURCES AND SYNTHESIS OF EVIDENCE; UPDATE TO EVIDENCE SYNTHESIS METHODS. REPORT BY THE NICE DECISION SUPPORT UNIT (April 2020).

*Recommendations*

- TSD-18 (Phillippo et al., 2016) advises on circumstances under which MAIC and STC could be used in submissions, and sets out some particulars of how they should be used and presented.
- We recommend that a new TSD is prepared to show how to use ML-NMR, along with worked examples and software code, and that the Methods Guide is revised to make it clear that MAICs should not be used under any circumstances, that STCs can be can be used for two-study scenarios, and that ML-NMR is the preferred approach for anchored comparisons. This could be developed over the next 6 months.

# Statistical performance

*Outcome modelling is perceived to perform better than weighting*

> Furthermore, MAICs perform poorly in simulation studies, and in some scenarios perform worse than standard NMA with no population adjustment (Phillippo, 2019).

> trial. Further details can be found in the key references, and a critique in TSD-18 (Phillippo et al., 2016). The same criticisms of MAIC apply to STC (see section 1.2.1), but with the exception that STC performs better in simulation studies than MAIC for the 2 study scenario (Phillippo, 2019).

> A recent simulation study (Phillippo, 2019) shows that ML-NMR performs similarly to STC in the 2-study scenario when the target population of interest is the population in the trial with aggregate data. However, it performs better than STC when the target population differs from that of the trial with aggregate data.

- Is MAIC biased in study *B*? ❌ If the target estimand is a conditional treatment effect, there will be bias because MAIC targets a marginal treatment effect. MAIC is unbiased if assumptions hold.

- Is MAIC potentially unprecise, therefore inefficient? ✅ Weighting methods have poor precision when the extremity of the weights is high and the effective sample size (ESS) after weighting is small.

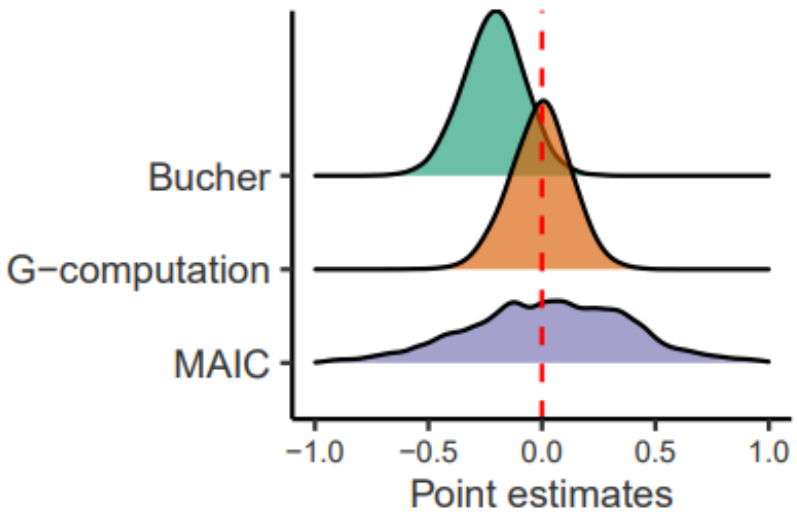
# Statistical performance

*Outcome modelling is perceived to perform better than weighting*

If assumptions for the methods hold, outcome modelling is more statistically precise and efficient than weighting, particularly if overlap is poor and/or the size of study *A* is small

Simulation study

- Anchored scenario, two RCTs with $N = 10000$, 1:1 randomization
- $X_k \sim \text{Normal}(0, 1)$ for study A and $X_k \sim \text{Normal}(-1.4, 1)$ for study B, $k = 1, 2, 3$ (poor overlap)
- $P(Y = 1 \mid X_1, X_2, X_3, T) = \text{expit}\big(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_T \mathbb{I}(T = active)\big)$
- $\beta_0 = -1$, $\beta_1 = \beta_2 = \beta_3 = 1$, $\beta_T = 1.05$
- MAIC balances the three covariate means and the outcome model is correctly specified




| Method | Bias | Coverage | MSE |
|---|---|---|---|
| Bucher | −0.211 (0.003) | 0.631 (0.011) | 0.062 (0.001) |
| G−computation | −0.006 (0.003) | 0.944 (0.005) | 0.018 (0.001) |
| MAIC | 0.034 (0.008) | 0.938 (0.005) | 0.137 (0.005) |

Remiro-Azócar, A., 2022. Purely prognostic variables may modify marginal treatment effects for non-collapsible effect measures. arXiv preprint arXiv:2210.01757.https://doi.org/10.48550/arXiv.2210.01757

# Weighting or outcome modelling?

*Beyond statistical precision and efficiency under no failures of assumptions*

## OUTCOME MODELLING

- Relies on model-based extrapolation to improve precision and efficiency
- Susceptible to bias when extrapolating a mis-specified outcome model
- Model misspecification bias difficult to detect; an outcome model that seems approximately correct in study A may not fit well in extrapolated regions
- Extrapolation uncertainty not accounted for
- Can produce the treatment effect estimates that are required for HTA where there is limited overlap

## WEIGHTING

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

# Standardizing with respect to "Study *B*"

*Considerations about marginal estimands*

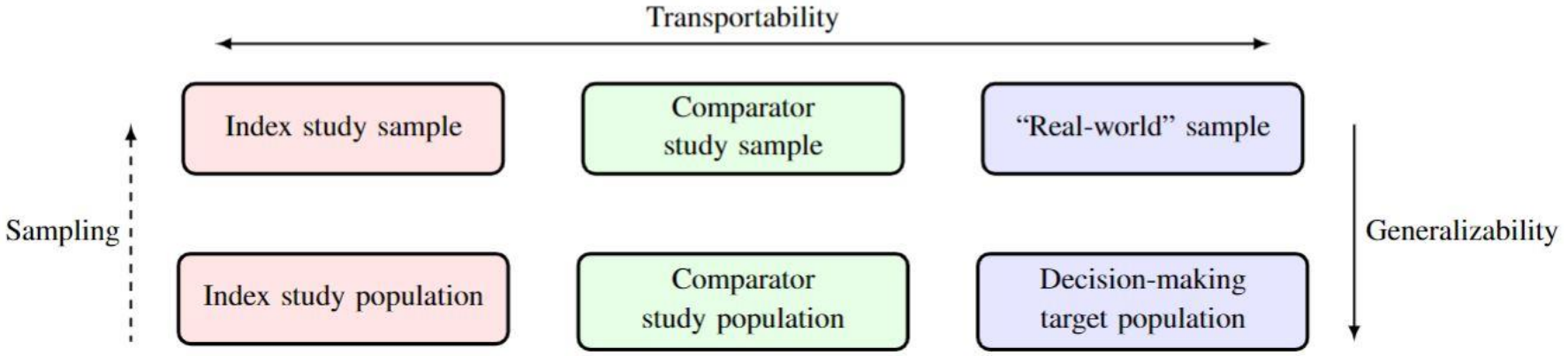
- Marginal estimands for non-collapsible measures, e.g. odds and hazard ratios:
  - Depend on the full distribution of measured and unmeasured covariates, not only on covariate means
  - May change with the distribution of "purely prognostic" covariates, i.e., variables that do not induce treatment effect heterogeneity at the individual level
  - Are not identifiable with limited access to patient-level data, without making further covariate distributional assumptions

- In our prior simulated example:
  - The outcome-generating model only contains main effects for the covariates and lacks treatment-covariate interactions; no effect measure modification for the (log) odds ratio at the individual level
  - Nevertheless, the marginal odds ratio for active treatment versus control is 2 in study *A* and 2.46 in study *B*

# Standardizing with respect to "Study *B*"

*Considerations about MAIC and STC*

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




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

- MAIC and STC are restricted to contrast treatments in the study *B* sample
- This may not be representative of the target population of eligible patients for study *B*
- Moreover, it may differ to the target population of routine clinical practice in the jurisdiction
- A valid estimate of the marginal effect in one context is not necessarily valid in another

In the anchored scenario, multilevel network meta-regression (ML-NMR) can potentially produce marginal effect estimates in any specified target population:

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

# Multilevel network meta-regression (ML-NMR)

**Define an individual-level regression model (IPD meta-regression)**

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

**Average (integrate) over the target population to form the aggregate-level model**

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

Phillippo, D.M., Dias, S., Ades, A.E., Belger, M., Brnabic, A., Schacht, A., Saure, D., Kadziola, Z. and Welton, N.J., 2020. Multilevel network meta-regression for population-adjusted treatment comparisons. Journal of the Royal Statistical Society. Series A,(Statistics in Society), 183(3), p.1189.

# Multilevel network meta-regression (ML-NMR)

- Parameterized on individual level-conditional treatment effects
- Conditional treatment effect at the covariate means

$$d_{ab(P)} = \int_{\boldsymbol{x}} \left(\mu_{(P)} + \boldsymbol{x}^{\mathsf{T}} \left(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,b}\right) + \gamma_b\right) f_{(P)}(\boldsymbol{x}) d\boldsymbol{x} - \int_{\boldsymbol{x}} \left(\mu_{(P)} + \boldsymbol{x}^{\mathsf{T}} \left(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2,a}\right) + \gamma_a\right) f_{(P)}(\boldsymbol{x}) d\boldsymbol{x}$$
$$= \bar{\boldsymbol{x}}_{(P)}^{\mathsf{T}} \left(\boldsymbol{\beta}_{2,b} - \boldsymbol{\beta}_{2,a}\right) + \gamma_b - \gamma_a,$$

- One can obtain marginal treatment effects through integration

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

$$\Delta_{ab(P)} = g(\bar{p}_{b(P)}) - g(\bar{p}_{a(P)}).$$

Phillippo, D., Dias, S., Ades, A.E. and Welton, N.J., 2021. Target estimands for efficient decision making: Response to comments on “Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study”. Statistics in Medicine, pp.2759-2763.

- **Open questions**
  - Application to disconnected networks (unanchored scenario)
  - Extension to survival analysis setting required

# Discussion: Implications for ITC/NMA in the context of health technology assessment