Bridging Single Arm Studies with Individual Participant Data in Network Meta-Analysis of Randomized Controlled Trials: A Simulation Study

Katerina Maria Kontouli[1], Stavros Nikolakopoulos[2], Christos Christogiannis[1], Dimitris Mavridis[1]

[1] Department of Primary Education, University of Ioannina, Greece
[2] Department of Psychology, University of Ioannina, Greece

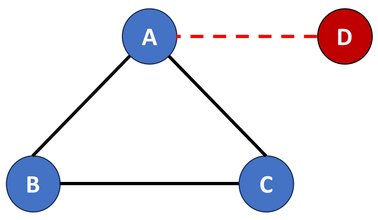
email: kmkontouli@uoi.gr | snikolakopoulos@uoi.gr | c.christogiannis@uoi.gr | dimi.mavridis@googlemail.com

Background

Why this matters: Health technology assessments (HTAs) increasingly consider evidence from **single-arm studies,** especially when randomized trials are limited or infeasible.

What is already known: The majority of published PAICs (56%) report significant results favoring the treatment from the IPD trial, while only 1 study (0.3%) favors the AgD comparator — highlighting strong evidence of publication bias and selective use of data to support sponsor-favored treatments.

The typical scenario: Manufacturers may have access to **individual participant data (IPD)** from their own single-arm study (e.g., treatment D), while only **aggregate data (AGD)** are available for other treatments (e.g., A, B, C).



The challenge: Conventional network meta-analysis (NMA) is designed for randomized comparisons and **does not directly accommodate non-randomized data**. Incorporating **single-arm studies** poses a risk of **bias due to the lack of randomization**, potential **confounding**, and differences in **baseline characteristics** between study populations.

Existing solutions: Methods like **Matching-Adjusted Indirect Comparison (MAIC)** and **Simulated Treatment Comparison (STC)** aim to adjust for **effect modifiers** and **prognostic factors**, enabling population adjusted indirect comparisons with RCTs.

Objectives

- To **evaluate and compare different statistical population adjustment methods —MAIC, STC, and Bucher Indirect Comparison**—for integrating single-arm studies into network meta-analysis.
- To understand **which method performs best under different conditions** using a simulation study.
- To provide **guidance on best practices for using single-arm evidence in decision-making**.

Methods

Simulation Design: We simulated a network including RCTs and a single-arm study with IPD. The target was to estimate indirect comparisons involving treatment D (from the single-arm study) against the other treatments (A, B, C) included in RCTs with AGD.

ADEMP	Simulation
Aims	Evaluate MAIC, STC, and Bucher for integrating single-arm studies into evidence synthesis.
Data-Generating Mechanism	Simulated data with a binary outcome and three continuous covariates. Apply distance metrics (Gower, Mahalanobis) for optimal matching.
Estimands	Relative treatment effect of the single arm vs the reference under different scenarios.
Methods	Implementation of MAIC, STC, and Bucher, with different matching strategies.
Performance Measures	Bias, MSE, confidence interval coverage

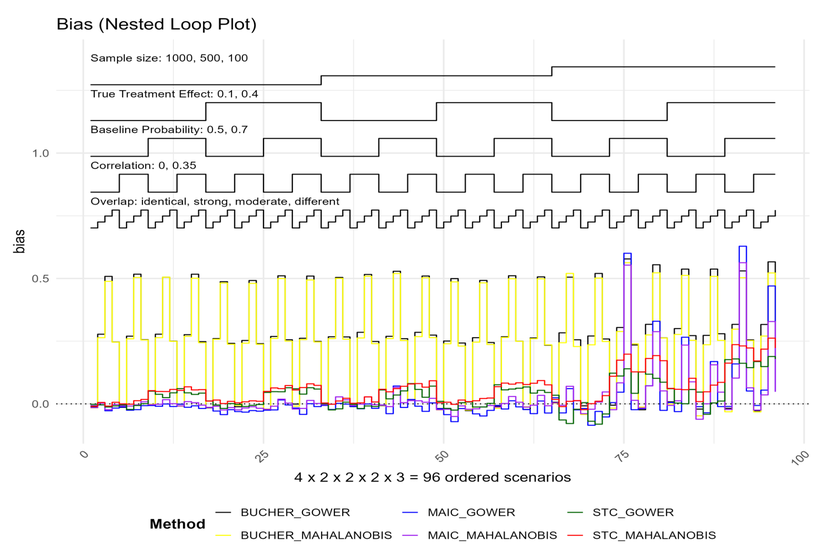
The scenarios:

1. **Varied the sample size of the studies** – (1000, 500, 100)
2. **Magnitude of Treatment Effect (TE)** – (Large: 0.40, Small: 0.10)
3. **Correlation of the three Covariates** – (No correlation, Moderate: 0.35)
4. **Baseline probability** – (0.5 or 0.7)
5. **Overlap Between Covariates** – (Identical overlap, Strong overlap, moderate overlap, different overlap)

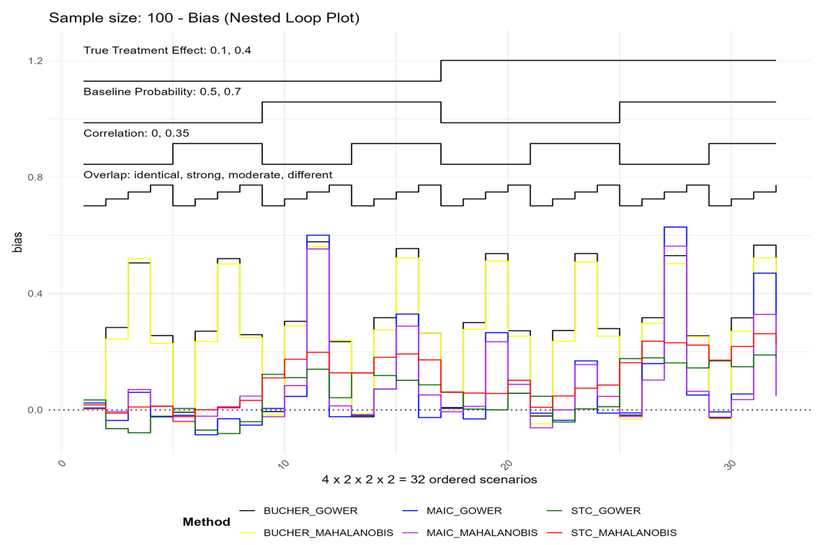
Number of scenarios: 3 x 2 x 2 x 2 x 4 =96

Results

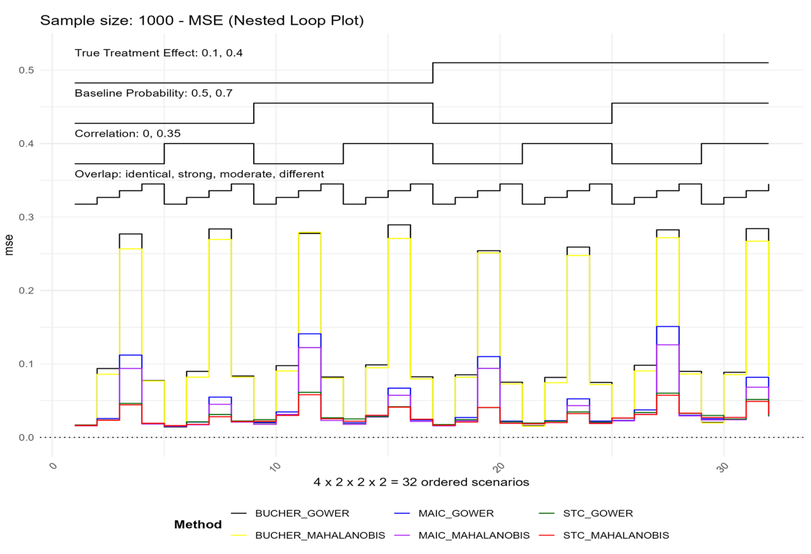
- **Bias of treatment effect estimates across 96 simulation scenarios**. STC (green and red lines) and MAIC (blue and purple lines) show consistently lower bias than the Bucher method (black and yellow lines), especially in scenarios with strong and moderate overlap.



- **Bias of treatment effect estimates across 32 simulation scenarios with sample size equal to 100.** STC consistently shows **lower bias** than MAIC, particularly in scenarios with lower covariate overlap (yellow bars). MAIC performance is more sensitive to overlap and covariate distribution.



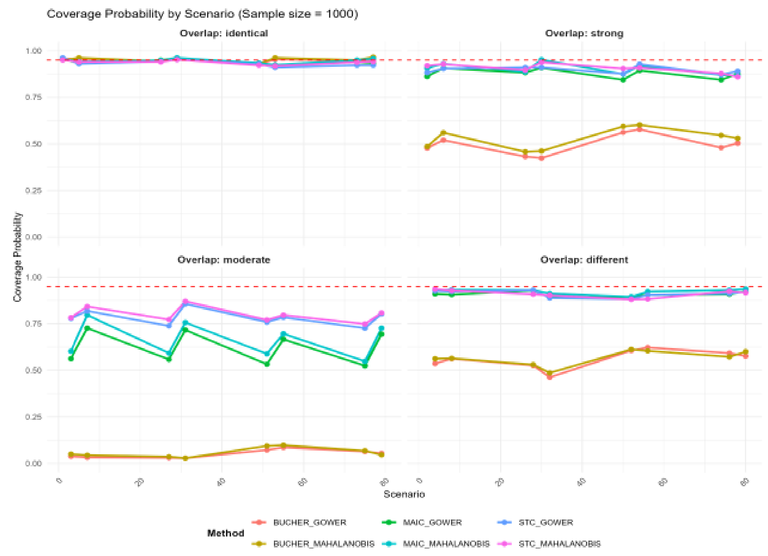
- **MSE across 32 scenarios with sample size equal to 1000.** With a larger sample size, **STC** consistently achieves the lowest MSE, especially under strong to identical overlap. **MAIC** also improves with sample size but shows higher variability. The **Bucher** method remains clearly inferior, while the **moderate-overlap scenarios** lead to increased MSE across all methods, emphasizing the importance of covariate balance



Contact information:

kmkontouli@gmail.com
kmkontouli@uot.gr

- **Coverage across 32 simulation scenarios with sample size equal to 1000.** STC consistently shows **better performance based on coverage** than MAIC and Bucher method. MAIC performance is more sensitive to overlap and covariate distribution.



Key Takeaways

- **No single gold standard method** — Continued research is essential.
- **STC methods** consistently show **better performance**, particularly for **MSE** and **coverage** across various scenarios
- **MAIC struggles** under **moderate overlap** conditions (around 60%)
- **Bucher's method is generally unreliable**, except in scenarios with **identical covariate distributions** and **large sample sizes**.
- **Mahalanobis matching** tends to outperform **Gower**, offering **more precise and consistent adjustments**

Conclusions & Next Steps

- **Quantify matching efficiency**: Evaluate the percentage of matched units using Gower vs. Mahalanobis distance.
- **Explore alternative distance metrics**
- **Expand to new scenarios**: Investigate performance under varying overlap levels, and correlation structures
- **Assess real-world data**: Apply these methods to real datasets to validate simulation findings.

References

1. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, Betts KA, Wu EQ. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health. 2012 Sep-Oct;15(6):940-7. doi: 10.1016/j.jval.2012.05.004. PMID: 22999145.
2. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. Med Decis Making. 2018 Feb;38(2):200-211. doi: 10.1177/0272989X17725740. Epub 2017 Aug 19. PMID: 28823204; PMCID: PMC5774635.
3. Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. Pharmacoeconomics. 2015 Jun;33(6):537-49. doi: 10.1007/s40273-015-0271-1. PMID: 25795232.
4. Ades AE, Welton NJ, Dias S, Phillippo DM, Caldwell DM. Twenty years of network meta-analysis: Continuing controversies and recent developments. Res Syn Meth. 2024; 15(5): 702-727. doi:10.1002/jrsm.1700.
5. Dimitris Mavridis - Network meta-analysis in a nutshell: Evidence Based Mental Health 2019;22:.
6. Gasparini, (2018). rsimsum: Summarise results from Monte Carlo simulation studies. Journal of Open Source Software, 3(26), 739, https://doi.org/10.21105/joss.00739
7. Serret-Larmande A, Zenati B, Dechartres A, Lambert J, Hajage D. A methodological review of population-adjusted indirect comparisons reveals inconsistent reporting and suggests publication bias. J Clin Epidemiol. 2023 Nov;163:1-10. doi: 10.1016/j.jclinepi.2023.09.004. Epub 2023 Sep 16. PMID: 37717707.

Acknowledgement

H.F.R.I. Hellenic Foundation for Research & Innovation
Greece 2.0 NATIONAL RECOVERY AND RESILIENCE PLAN
EVIDENCE SYNTHESIS METHODS STATISTICS TEAM
UNIVERSITY of IOANNINA