Stochastic Subset-Selection Method for Anchored Population-Adjusted Indirect Treatment Comparisons: SWARM-ITC

AUTHORS: Amit Kumar, BSc[1], Mahendra Kumar Rai, PhD[2], Deepthi Chinchapattanam, M.Pharm[1], Prachi Singh, MS (Pharm)[2], Manpreet Singh Kalsey, MBA[1], **Manasa Vishnubhotla, MS[1]**

[1] Trinity Life Sciences, Mumbai, India; [2] Trinity Life Sciences, Singapore.

MSR23

Poster presented at ISPOR AP Summit 2026

Manasa Vishnubhotla | mvishnubhotla@trinitypartners.com

TRINITY

1 Background

- Indirect treatment comparison (ITC) is an indispensable comparative effectiveness methodology for new treatments that usually have limited or insufficient comparative data for informed healthcare decision-making.[1,2]
- Early ITC approaches were limited by cross-trial differences in prognostic or effect-modifying patient characteristics, and by use of aggregate-level data (ALD) only, which precludes standard adjustment techniques.[3,4,5]
- Matching-adjusted indirect comparison (MAIC) addressed these limitations and is reported to be widely used for health technology assessment (HTA) submissions.[5,6]
- Proposed MAIC modifications (entropy balancing, alternative weighting, Poly MAIC,[7] 2 Staged MAIC [8] can improve precision and effective sample size (ESS), but unobserved differences may remain.

EVIDENCE GAP In the absence of head-to-head comparative evidence, generating new trial data is often impractical due to cost and time constraints, making ITCs a necessary alternative. MAIC is commonly used to adjust for cross-trial differences and, for a given set of matching moments, yields a single method-of-moments weighting solution. Its recognised limitations lie elsewhere: sensitivity to the choice of matched population and matching moments, loss of effective sample size under extreme weights, and finite-sample uncertainty in which patients effectively contribute to the comparison[7,10]. Conventional MAIC does not quantify this uncertainty in the choice of matched population, potentially leaving an important source of uncertainty unaccounted for in ITC.

Current evidence context: ITC / MAIC widely used in HTA → **Evidence gap: uncertainty in the matched population not captured** → **Need for an alternative population-adjusted ITC**

2 Objective

To develop SWARM-ITC, a stochastic subset-selection method for anchored population-adjusted indirect treatment comparison, and to evaluate it as a proof of concept against conventional MAIC using simulated randomized controlled trial datasets, comparing the two methods on accuracy of baseline matching, estimated treatment effects (hazard ratio and median progression-free survival), the associated uncertainty, and stability of estimates across two reciprocal scenarios.

3 Methods

In the absence of suitable patient-level comparative datasets for proof-of-concept evaluation, realistic simulated randomized controlled trial (RCT) datasets were generated to compare SWARM-ITC with MAIC in an anchored ITC setting.

3.1 Simulated (synthetic) data generation

- Two RCT datasets were simulated in R with the simstudy package,[9] which first defines the data elements and then generates the corresponding patient records.
- Five baseline covariates reflecting an oncology indication were specified: age > 60 years, gender, prior stem cell transplant (SCT), prior lines of treatment > 3 (PRLINE>3), and International Staging System (ISS) stage I, II or III.
- Patients were then generated under 1:1 allocation, with 300 per arm in RCT1 (arm A vs B) and 250 per arm in RCT2 (arm B vs C), using comparable covariate distributions across both trials.
- Time-to-progression or death was finally simulated for each trial with the simsurv package,[9] drawing survival times from a two-parameter Weibull model (shape, scale) conditional on the covariates above.

Comparison of interest: A vs C (no direct trial)

RCT1
Simulated 1:1 randomized trial, total n = 600
Treatment A n = 300 | **Treatment B n = 300**

Arm B appears in both trials, so it anchors the comparison.
Treatment B Is the common comparator

RCT2
Simulated 1:1 randomized trial, total n = 500
Treatment B n = 250 | **Treatment C n = 250**

3.2 MAIC methodology

- Relevant prognostic factors and effect modifiers are selected first, and the baseline characteristics reported for the comparator trial are set as the matching target.
- As a proof of concept, two reciprocal scenarios were evaluated: Scenario 1 used IPD from RCT1 adjusted to the aggregate data of RCT2, and Scenario 2 used IPD from RCT2 adjusted to the aggregate data of RCT1.
- Following Signorovitch et al., 2010,[5] IPD in each scenario are reweighted so that their baseline characteristics (mean, median, standard deviation) match the aggregate values of the target trial.
- Outcomes in both arms of the IPD trial are re-estimated using these weights, and the adjusted effect is anchored through the common comparator arm B (Bucher method)[1] to obtain the indirect A vs C comparison in each scenario.

3.3 SWARM-ITC approach

- The same baseline characteristics selected for matching are used, with the aggregate values of the target trial defining the matching target, applied to both reciprocal scenarios.
- In each iteration, a stochastic combinatorial optimization algorithm selects a subset of patient records from the trial with IPD (RCT1: A vs B) so that the subset reproduces the aggregate baseline characteristics (mean, median, standard deviation) of the comparator arm of the target trial (RCT2: B vs C), with no weighting involved.
- Matching accuracy is verified before outcome analysis, and the selection and outcome analysis are repeated over 1,000 iterations, each yielding one balanced subset and one treatment effect.
- Each adjusted effect is anchored through the common comparator arm B (Bucher method)[1], and the distribution of effects across iterations provides the median and credible interval for the A vs C comparison; stability was assessed using trace plots.

Four steps of the SWARM-ITC methodology

Step 1	Selection of baseline characteristics for matching (prognostic factors and effect modifiers), agreed with clinical and statistical experts
Step 2	Assessment of baseline similarity between the IPD trial (RCT1) and the aggregate comparator trial (RCT2) to confirm feasibility of matching
Step 3	Selection of a balanced patient subset by combinatorial optimization, under constraints on arm size, covariate balance and patient retention
Step 4	Verification of matching accuracy against the aggregate targets of the comparator trial, performed before outcome analysis in each iteration

Each iteration yields one balanced subset, and one corresponding treatment effect estimate with its own within-subset variance. Repeating the procedure across iterations generates a distribution of estimates, summarized as the median with a credible interval derived by pooling iteration-level variance using Rubin's rules[11], so that within-subset sampling error and between-iteration variability from the stochastic matching solution are both propagated rather than assumed away by a single set of weights.

3.4 Scenarios

- **Scenario 1:** individual patient data of RCT1 and aggregate-level data of RCT2.
- **Scenario 2:** individual patient data of RCT2 and aggregate-level data of RCT1, to evaluate consistency of results across both scenarios.

3.5 SWARM-ITC mathematical model

Sets and parameters

- SWARM-ITC is formulated as a binary optimization problem, specified by decision variables, model parameters, constraints and an objective function.
- Indices: i ∈ (1…i) patients in the IPD; j ∈ (1…j) treatment arms in the IPD; k ∈ (1…k) covariates selected for matching.
- Decision variable: X_{ij} = 1 if patient i in arm j is selected, 0 otherwise.
- Parameters: IPD_{ijk}, the k^{th} baseline characteristic of patient i in arm j; B_k, the k^{th} baseline characteristic reported in the published aggregate data; θ, the proportion of patients excluded; N, the number of IPD observations available at the start of the analysis.

Constraint 1: Equal number of selected patients per arm

$$\sum_i \sum_{j=1} X_{ij} = \sum_i \sum_{j=2} X_{ij} \quad \text{(equation 1)}$$

Constraint 2: Balanced covariate distribution across arms

$$\frac{\sum_i \sum_{j=1} X_{ij} IPD_{ijk}}{\sum_i \sum_{j=1} X_{ij}} = \frac{\sum_i \sum_{j=2} X_{ij} IPD_{ijk}}{\sum_i \sum_{j=2} X_{ij}} \quad \text{(equation 2)}$$

Constraint 3: Maximum retention of available patients

$$\sum_i \sum_j X_{ij} > N(1-\theta) \quad \text{(equation 3)}$$

Objective: Minimize deviation from aggregate data

$$\sum_K \left[\sum_i \sum_j IPD_{ijk}\ X_{ij} / \sum_i \sum_j X_{ij} - B_k\right] 2 \quad \text{(equation 4)}$$

- Equations 1 to 4 were solved in R using a stochastic binary particle swarm optimization algorithm[12]; a V-shaped transfer function[13] mapped particle velocities onto the binary decision variables X_{ij}, so each run returns one feasible subset satisfying all constraints.
- Interval derivation: across M = 1,000 iterations, the log-hazard ratio and its within-subset variance were pooled by Rubin's rules[11].
- Total variance T = W + (1 + 1/M)B, where W is the mean within-iteration variance and B the between-iteration variance.
- Credible interval = median ± 1.96√T, back-transformed to the hazard ratio scale, propagating both within-subset sampling error and between-subset variability.

3 Methods

3.6 MAIC vs SWARM-ITC

Attribute	MAIC	SWARM-ITC
Adjustment approach	Reweights IPD from the trial with patient level data so that weighted baseline characteristics match those reported for the target trial (Signorovitch et al., 2010; NICE DSU TSD 18) [5,10]	Selects, by binary combinatorial optimization, a subset of IPD records whose baseline characteristics match the aggregate values reported for the target trial
Balancing mechanism	Method of moments weights of propensity score form; requires inspection of the weight distribution and reporting of the effective sample size (ESS)	Patient selection under explicit constraints; no weights are estimated
Nature of algorithm	Deterministic: one weight solution is obtained for a given set of matching variables	Stochastic: repeated optimization yields multiple admissible subsets; higher computational requirement
Balance and precision	Exact on the matched moments, but extreme weights can substantially reduce ESS and precision	Constraints enforce covariate balance and equal numbers per arm; analysis is based on a retained subset of the IPD
Summary measures matched	Typically means and standard deviations, and proportions for categorical variables, as reported in the target trial	Means, medians and standard deviations; extendable to further moments or covariate correlations by modifying the constraints or objective function
Uncertainty and output	Point estimate with confidence interval, based on robust or bootstrap standard errors	Distribution of estimates across iterations, summarized as median with credible interval; within- and between-iteration variance pooled by Rubin's rules; convergence assessed using trace plots
Assumptions in this analysis	Anchored comparison via the common comparator; assumes constancy of relative treatment effects and that all effect modifiers are reported	Anchored comparison under the same assumptions, with balance achieved by patient selection rather than weighting

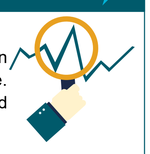
3.7 Methods Overview

Method component	Description	Details	Role in analysis
Evidence base	Two simulated randomized controlled trials	RCT1 (A vs B) and RCT2 (B vs C), anchored through the common comparator arm B	Provides the indirect comparison of interest, A vs C
Synthetic data generation	Simulated baseline characteristics and outcome data	simstudy R package [9]; five covariates (age > 60 years, gender, prior SCT, PRLINE>3, ISS stage); 300 per arm (RCT1), 250 per arm (RCT2)	Covariate dependencies imposed: prior SCT as a function of age, PRLINE>3 as a function of prior SCT
SWARM-ITC approach	Stochastic subset selection by binary combinatorial optimization	Each iteration selects an IPD subset matching the aggregate baseline characteristics of the comparator trial; 1,000 iterations generated distinct admissible subsets	Matching accuracy verified in each iteration before outcome analysis
MAIC approach	Reweighting of IPD to the aggregate comparator data	Method of moments weights (Signorovitch et al. 2010 approach) [5] estimated by matching IPD baseline characteristics to the aggregate target trial, then applied to the survival analysis	Serves as the established comparator method for anchored adjustment
MAIC vs SWARM-ITC comparison	Head-to-head methodological comparison on the same simulated data	Two reciprocal scenario analyses assessing matching accuracy, pooled treatment effects and stability	Summarised in the MAIC vs SWARM-ITC panel
Outcomes analyzed	Time-to-event outcomes	Survival time simulated from a two-parameter Weibull model (simsurv) [9]; time to progression; hazard ratio and median PFS reported	Adjusted effects combined across trials using the Bucher method[1]
Covariates adjusted	Baseline characteristics selected for matching	Age > 60 years, % male, prior SCT, PRLINE>3, ISS stage I / II / III	Prognostic factors and effect modifiers agreed with clinical and statistical experts
Uncertainty / sensitivity analysis	Scenario and stability assessment	Scenario 1: RCT1 IPD with RCT2 aggregate data; Scenario 2: RCT2 IPD with RCT1 aggregate data; stability assessed by trace plots at 100, 500 and 900 iterations	Uncertainty expressed as credible intervals (SWARM-ITC; iteration-level variance pooled via Rubin's rules) and confidence intervals (MAIC)

4 Results

Accuracy of matching

Assessment of baseline comparability demonstrated that both MAIC and SWARM-ITC closely aligned with the target comparator population across key prognostic and effect-modifying variables, including sex, age >60 years, prior SCT, prior lines of therapy, and ISS stage. Differences between adjusted and target population characteristics were ≤1 percentage point, indicating negligible residual imbalance and supporting the validity of the anchored comparative effectiveness estimates.



Scenario 1: RCT1 IPD vs RCT 2 aggregate data

Table 1. Accuracy of matching (RCT1 IPD vs RCT 2)

Summary baseline characteristics	RCT1 (A vs B): Unadjusted			RCT2 (B vs C)			RCT1: MAIC adjusted			RCT1: SWARM-ITC adjusted		
	Overall	Arm A	Arm B	Overall	Arm B	Arm C	Overall	Arm A	Arm B	Overall	Arm A	Arm B
Sample size /ESS*	600	300	300	500	250	250	466	233	233	304	152	152
% male	41.5%	41.7%	41.3%	51.8%	51.2%	52.4%	51.8%	51.9%	51.7%	51.3%	51.3%	51.3%
Age > 60 yrs.	42.7%	42.7%	42.7%	43.4%	43.2%	43.6%	43.4%	43.4%	43.4%	43.4%	43.4%	43.4%
Prior SCT	26.5%	27.0%	26.0%	27.8%	28.0%	27.6%	27.8%	28.1%	27.5%	27.6%	27.6%	27.6%
PRLINE>3	36.2%	36.0%	36.3%	34.8%	35.2%	34.4%	34.8%	34.6%	35.0%	34.9%	34.9%	34.9%
ISS I	22.8%	23.0%	22.7%	42.4%	42.4%	42.4%	42.4%	42.5%	42.3%	42.1%	42.1%	42.1%
ISS II	34.5%	34.3%	34.7%	28.2%	28.4%	28.0%	28.2%	28.1%	28.3%	28.3%	28.3%	28.3%
ISS III	42.7%	42.7%	42.7%	29.4%	29.2%	29.6%	29.4%	29.4%	29.4%	29.6%	29.6%	29.6%

Abbreviation: ESS, effective sample size; ISS, international staging system; PRLINE>3, prior line of treatment greater than 3; SCT, stem cell transplant

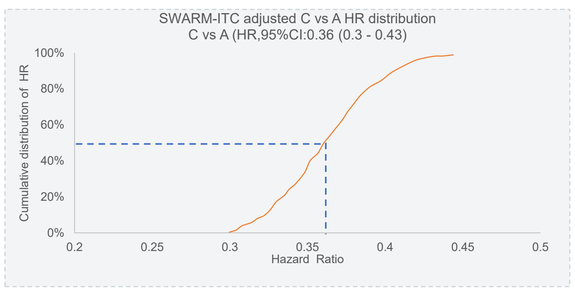
Scenario 2: RCT2 IPD vs RCT 1 aggregate data

Table 2. Accuracy of matching (RCT2 IPD vs RCT 1)

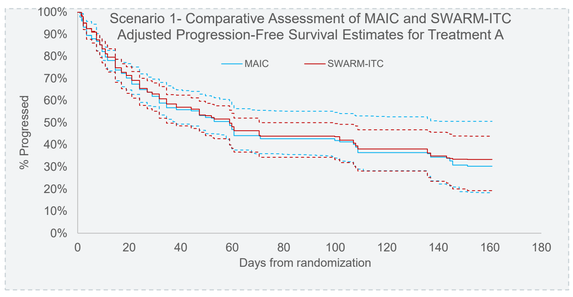
Summary baseline characteristics	RCT2 (B vs C) Unadjusted			RCT1 (A vs B)			RCT2: MAIC adjusted			RCT2: SWARM-ITC adjusted		
	Overall	Arm B	Arm C	Overall	Arm A	Arm B	Overall	Arm B	Arm C	Overall	Arm B	Arm C
Sample size /ESS*	500	250	250	600	300	300	415	207	208	254	127	127
% male	51.8%	51.2%	52.4%	41.5%	41.7%	41.3%	41.5%	41.0%	42.0%	41.7%	41.7%	41.7%
Age > 60 yrs.	43.4%	43.2%	43.6%	42.7%	42.7%	42.7%	42.7%	42.8%	42.6%	42.5%	42.5%	42.5%
Prior SCT	27.8%	28.0%	27.6%	26.5%	27.0%	26.0%	26.5%	26.8%	26.2%	26.8%	26.8%	26.8%
PRLINE>3	34.8%	35.2%	34.4%	36.2%	36.0%	36.3%	36.2%	37.0%	35.3%	36.2%	36.2%	36.2%
ISS I	42.4%	42.4%	42.4%	22.8%	23.0%	22.7%	22.8%	22.9%	22.8%	22.8%	22.8%	22.8%
ISS II	28.2%	28.4%	28.0%	34.5%	34.3%	34.7%	34.5%	34.8%	34.2%	34.6%	34.6%	34.6%
ISS III	29.4%	29.2%	29.6%	42.7%	42.7%	42.7%	42.7%	42.4%	43.0%	42.5%	42.5%	42.5%

Abbreviation: ESS, effective sample size; ISS, international staging system; PRLINE>3, prior line of treatment greater than 3; SCT, stem cell transplant

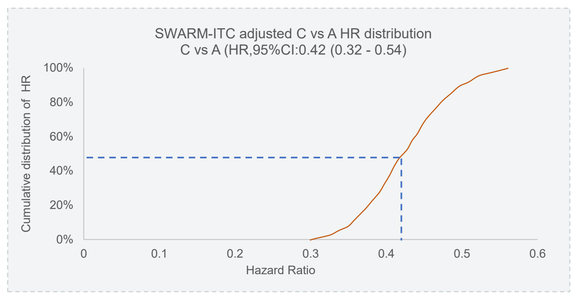
Scenario 1: SWARM-ITC HR Distribution Across 1000 iterations (C vs A)



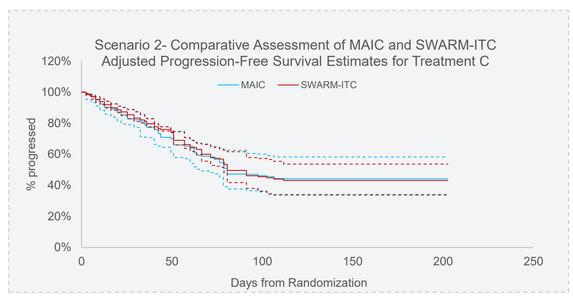
Scenario 1 - Adjusted PFS for drug A: MAIC vs SWARM-ITC



Scenario 2: SWARM-ITC HR Distribution Across 1000 iterations (C vs A)



Scenario 2- Adjusted PFS for drug C: MAIC vs SWARM-ITC



Scenario 1 | Anchored adjusted estimates

Hazard ratio (C vs A)	Median PFS, treatment A (days)
MAIC 0.37 (0.16–0.84)	MAIC 58 d (32–134)
SWARM-ITC 0.36 (0.30–0.43)	SWARM-ITC 52 d (37–59)
0.2 1.0	0 100 200

Scenario 2 | Anchored adjusted estimates

Hazard ratio (C vs A)	Median PFS, treatment C (days)
MAIC 0.43 (0.19–0.96)	MAIC 79 d (63–NR)
SWARM-ITC 0.42 (0.32–0.54)	SWARM-ITC 78 d (63–NR)
0.2 1.0	0 100 200

Stability of results

Trace plots showed results stabilized as iterations increased from 100 to 500 and eventually 900 for median, upper and lower credible interval (almost straight lines at 900).

5 Limitations

- Findings are based on simulated data and a limited set of scenarios, so they may not reflect real-world practice and cannot be generalized.
- Outcomes assumed a Weibull model with log-linear covariate effects; non-linear or non-proportional hazard settings were not assessed.
- Matching is possible only for characteristics reported at the aggregate level, and validity relies on constancy of relative treatment effects across trials.
- Matching on non-central moments was not evaluated, so performance against the Signorovitch approach[5] could not be compared for these measures.
- SWARM-ITC is stochastic and requires greater computational resource and time than the deterministic MAIC approach.

6 Conclusions

SWARM-ITC provided a feasible alternative to MAIC for anchored population-adjusted indirect comparison.

Adjusted baseline characteristics were closely aligned between methods, balance being achieved by patient selection rather than weighting.

SWARM-ITC reproduced MAIC point estimates with comparable baseline balance. Narrower intervals should not be read as greater efficiency, as fewer patients are retained than the MAIC effective sample size; the pooled variance components require formal evaluation.

Further scenarios, real evidence settings and algorithm refinement to reduce computation time are warranted.

FUNDING & DISCLOSURE

- This study was conducted and internally funded by Trinity Life Sciences. All authors are employees of Trinity Life Sciences. The analysis used simulated datasets only; no patient-level data were used, and no ethics approval was required. The authors have no other conflicts of interest to declare.

REFERENCES

1. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997; 50: 683-91.
2. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med. 2009; 151: 203-5.
3. Alsop JC, Pont LO. Matching-adjusted indirect comparison via a polynomial-based non-linear optimization method. J Comp Eff Res. 2022; 11: 551-61.
4. Petto H, Kadziola Z, Brnabic A, et al. Alternative weighting approaches for anchored matching-adjusted indirect comparisons via a common comparator. Value in Health. 2019; 22: 85-91.
5. Signorovitch JE, Wu EQ, Yu AP, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. Pharmacoeconomics. 2010; 28: 935-45.
6. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health. 2012; 15: 940-7.
7. Remiro-Azócar A, et al. A framework for reliable, transparent, and reproducible population-adjusted indirect comparisons. Pharmacoeconomics. 2025
8. Remiro-Azócar A. Two-stage matching-adjusted indirect comparison. BMC Med Res Methodol. 2022;22:223. doi:10.1186/s12874-022-01692-9.
9. Goldfeld K, Wujciak-Jens J. simstudy: Illuminating research methods through data generation. J Open-Source Softw. 2020;5(54):2763
10. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE. Sheffield, UK: NICE Decision Support Unit; 2016.
11. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons; 1987.
12. Kennedy J, Eberhart RC. A discrete binary version of the particle swarm algorithm. Proc IEEE Int Conf Syst Man Cybern. 1997; 5: 4104-8.
13. Mirjalili S, Lewis A. S-shaped versus V-shaped transfer functions for binary particle swarm optimization. Swarm Evol Comput. 2013; 9: 1-14.

CONTACT

- Amit Kumar, Trinity Life Sciences, Mumbai, India | amkumar@trinitypartners.com