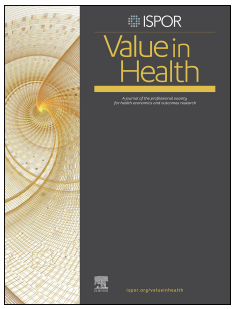
A comparison of relative-efficacy estimate(s) derived from both matching-adjusted indirect comparisons and standard anchored indirect treatment comparisons

Journal Pre-proof



A comparison of relative-efficacy estimate(s) derived from both Matching-Adjusted Indirect Comparisons and Standard Anchored Indirect Treatment Comparisons

Owen Cassidy, MSc, Marie Harte, MSc, Lea Trela-Larsen, PhD, Cathal Walsh, PhD, Arthur White, PhD, Laura McCullagh, PhD, Joy Leahy, PhD

PII: S1098-3015(23)03061-9

DOI: https://doi.org/10.1016/j.jval.2023.07.001

Reference: JVAL 3844

To appear in: *Value in Health*

Received Date: 1 November 2022

Revised Date: 6 June 2023

Accepted Date: 7 July 2023

Please cite this article as: Cassidy O, Harte M, Trela-Larsen L, Walsh C, White A, McCullagh L, Leahy J, A comparison of relative-efficacy estimate(s) derived from both Matching-Adjusted Indirect Comparisons and Standard Anchored Indirect Treatment Comparisons, *Value in Health* (2023), doi: https://doi.org/10.1016/j.jval.2023.07.001.

Target Journal: VIH

Title: A comparison of relative-efficacy estimate(s) derived from both Matching-Adjusted Indirect Comparisons and Standard Anchored Indirect Treatment Comparisons
Subtitle: A Review of Matching-Adjusted Indirect Comparisons

Authors: Owen Cassidy MSc[1,2], Marie Harte MSc[1,2], Lea Trela-Larsen PhD[1,2], Cathal Walsh PhD [3], Arthur White PhD [4], Laura McCullagh PhD [1,2] and Joy Leahy PhD [1,2]

Affiliations: 1. National Centre for Pharmacoeconomics Ireland, St. James's Hospital, Dublin, Ireland; 2. Department of Pharmacology and Therapeutics, Trinity College Dublin, Dublin, Ireland; 3. Health Research Institute and MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland 4. Department of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland.

Corresponding Author information:

Joy Leahy, PhD

National Centre for Pharmacoeconomics Ireland

St. James's Hospital; Trinity College Dublin

Dublin, Ireland.

Email: jleahy@stjames.ie

Phone: +35314103427

Precis: We analyse real world Matching-Adjusted Indirect Comparisons to ascertain if they tend to favour the treatment for which individual patient data is available.

Word count: 3945
Number of pages: 17
Number of figures: 4
Number of tables: 2

Supplementary materials:

Pages: Excel file

Figures: None

Tables: 1 (excel file)

Author contributions:

Concept and design: Leahy, Harte, White
Acquisition of data: Harte
Analysis and interpretation of data: Cassidy, Trela-Larsen, Walsh, White, Leahy
Drafting of the manuscript: Cassidy, Harte, Walsh, McCullagh, Leahy,
Critical revision of the paper for important intellectual content: Harte, Trela-Larsen, Walsh, White, McCullagh, Leahy
Statistical analysis: Cassidy, White, Leahy
Obtaining funding: None
Administrative, technical, or logistic support: None
Supervision: Trela-Larsen, McCullagh, Leahy

Conflict of Interest Disclosures: The authors report no conflicts of interest.

Role of the Funders/Sponsors: The authors received no financial support for this research.

Acknowledgments: None

Highlights

- Matching-Adjusted Indirect Comparisons (MAIC) is a method which is increasing in popularity in evidence synthesis. A number of simulation studies have been carried out to assess its suitability as a method. However, simulation studies alone cannot assess bias for MAIC given the subjective nature of how or if at all MAIC is performed.
- We analyse real world MAICs to ascertain if they tend to favour the treatment for which individual patient data is available, relative to the results of the standard anchored indirect treatment comparisons (ITCs).
- This work provides an insight into how effect estimates compare when MAIC results are reported or not reported. This can be useful for analysts to consider when appraising results arising from MAIC, or when evaluating the absence of a reported MAIC.

Abstract

Objectives: We present an empirical comparison of relative-efficacy estimate(s) from Matching-Adjusted Indirect Comparisons (MAICs) with estimates from corresponding standard anchored indirect treatment comparisons (ITCs).

Methods: A total of 80 comparisons were identified from 17 publications through a systematic rapid review. A standardised metric that used reported relative treatment efficacy estimates and their associated uncertainty was used to compare the methods across different treatment indications and outcome measures.

Results: On aggregate, MAICs presented for connected networks tended to report a more favourable relative-efficacy estimate for the treatment for which individual-level patient data (IPD) were available, relative to the reported ITC estimate.

Conclusions: Although we recognise the importance of MAIC and other population adjustment methods in certain situations, we recommend that results from these analyses are interpreted with caution. Researchers and analysts should carefully consider if MAICs are appropriate where presented and whether MAICs would have added value where omitted.

Introduction

Comparative-effectiveness research (sometimes called relative-effectiveness research) is routinely performed in research and development of pharmaceutical interventions. While several competing treatments may exist for a given disease state, there may be no multi-arm study available to directly compare all relevant treatments. In the absence of direct evidence, investigators can perform indirect treatment comparisons (ITC) using estimates from clinical trials [1].

Where a common comparator exists, a standard anchored ITC (also known as a special case of Network Meta-Analysis (NMA)) may be performed, as introduced by Bucher et al [2]. A standard anchored ITC compares the relative effects of two treatments, versus a common comparator, to obtain an estimate for the comparison of interest. This method has been widely adopted within comparative-effectiveness research and is recognised within reimbursement bodies [3-6].

All ITC evidence may be subject to bias. Lunny et al [7] have identified 22 sources of bias specific to ITC. Using indirect evidence to estimate relative treatment effects often requires strong assumptions regarding the exchangeability of evidence. There can also be different modelling options. Whenever multiple modelling options exist the analyst must consider the impact of choosing one approach over another. It may be difficult to judge which approach is most appropriate. For example, an analyst will need to judge whether a standard anchored ITC is more appropriate than more complex population adjustment methods.

We compare the evidence-synthesis outputs of published MAICs (identified via rapid review) with the outputs of the corresponding standard anchored ITC. We investigate whether more favourable relative efficacy estimates (for the intervention under consideration versus the comparator(s)) tend to be produced for one method. For this work, a corresponding standard

anchored ITC is one that is informed by the same clinical trials as the MAIC under consideration, but without patient-level adjustment. We do not aim to guide analysts on the appropriateness of performing a particular MAIC. However, we provide our observations of MAIC, as applied in the literature, which may be useful for HTA analysts and appraisers to consider, as well as other researchers.

MAIC [8] has been introduced to perform the comparison within the same two-study scenario as a standard anchored ITC. This is referred to as an anchored MAIC. In situations where no common comparator exists, the unanchored MAIC has been proposed. However, we focus on the anchored scenario in which the analyst can choose to perform both an MAIC and a standard anchored ITC. Thus, we can compare the evidence synthesis outputs for both methods.

Results from standard anchored ITCs may be biased when certain patient characteristics, known as effect modifiers, are imbalanced between two trials [9]. For effect modifiers, a change in the patient characteristics can lead to differing changes in treatment efficacy among treatments. Thus, bias may be introduced in the standard anchored ITC result. In this case, one of the fundamental assumptions of ITC may not hold, i.e. that trials are not exchangeable or the patient population is not sufficiently homogenous. However, if individual patient data (IPD) is available, then matching methods such as MAIC, or related methods such as propensity scoring can be useful in adjusting for these effect modifiers. The patient populations then compared are likely to have improved homogeneity for these effect modifiers, thus reducing this bias.

MAIC requires access to IPD for one of the trials comparing an intervention. In theory, this strengthens the comparison by accounting for some of the difference in relative effects induced by differences in the trial populations. However, IPD is generally only provided to the analyst at the discretion of the pharmaceutical company. We refer to the intervention for which IPD are available as the "IPD treatment". The competing intervention, for which aggregate data only is available, is referred to as the "AgD treatment" and the common comparator forming the indirect comparison is referred to as the "anchor treatment". This schematic is outlined in Figure 1. MAIC is based on propensity scoring methodology. However, a limitation of MAIC is that IPD is only available from one of the trials.

Many simulation studies have assessed the performance of MAIC. However, many of these studies yielded differing results. Some differing results are due to the exact nature of the question being posed when undertaking an MAIC and in which population of interest. Phillipo et al [10] suggested that MAICs may increase bias compared with standard ITCs. However, these results have been challenged [11], as they point out that MAIC targets a marginal or population-average treatment effect rather than a conditional treatment effect. Additionally, theoretical results by Cheng et al [12, 13] suggest that MAIC may be more bias-robust than STC. However, STC, when adapted to target a marginal effect, as demonstrated in Remiro Azocar et al [15], can produce more precise and accurate estimates than MAIC, but relies on correct specification of the outcome model. . Another simulation study by Leahy et al [16] found that MAICs may often be of limited benefit in anchored comparisons, but can be useful as a sensitivity analysis. However, this finding only applies when the population to which the IPD is being reweighted is representative of a target population of interest. To date,

current conclusions are based on simulation studies and no empirical analysis of MAIC methods has been performed. Due to the subjective nature of decisions in patient characteristic selection in MAIC, and even whether to perform or publish an MAIC, we undertake a comparison of the evidence-synthesis outputs of published MAICs with the outputs of the corresponding standard anchored ITC. This work will be useful to inform HTA analysts and assessors in selecting methods or combinations of methods based on their operating characteristics, and the particular data and research questions at hand.

Methods

Identification of MAICs and corresponding, standard anchored ITCs.

Due to the documented resource burden of systematic reviews, we utilised an accelerated systematic review process (a rapid review) to identify published anchored MAIC studies [17]. A rapid review is a type of knowledge synthesis in which components of the traditional systematic review process are simplified or omitted to produce evidence in a more efficient manner [17,18]. Rapid reviews are therefore an effective tool which incorporates transparent, scientific, and reproducible methodology [19]. On the 25 February 2019 searches were undertaken using three databases: Embase, Medline (via OVID), and the Cochrane Library. Grey literature was also searched using Grey Matters Light tool. Citation mining (backwards and forwards) was also undertaken for all journal articles included in the final analysis. For this review, the screening and data-extraction stages of a traditional systematic literature review were simplified. Both title and abstract screening and full text screening were undertaken by a single reviewer. Data extraction was also undertaken by a single reviewer and validated by a second member of the review team.

For each identified MAIC, corresponding standard anchored ITCs were identified. In certain instances, both an MAIC and a corresponding standard anchored ITC were reported within the same publication. Where a corresponding standard anchored ITC was not reported in the same publication, published corresponding ITCs were identified via targeted literature searches.

In some cases, multiple publications referred to the same primary studies. In other cases, multiple outcomes were reported within a publication. To accurately account for this a strict protocol for the analysis was established.

Protocol

The number of distinct treatment comparisons is determined from the set of screened records. Multiple outcomes within a record are referred to as subrecords. Therefore, each individual citation corresponds to one record, whereas a subrecord is one of many evidence synthesis outcomes within one citation. Where multiple records refer to the same treatment comparison, the procedure detailed in Table 1 was applied.

We obtained the point estimates and confidence/credible interval (CI/CrI) of relative efficacy from all the identified MAICs and the corresponding standard anchored ITCs.

Data extraction and Analysis

As many different outcomes were included, the absolute value of the raw difference between the MAIC and the ITC results varied considerably between subrecords. For example, in numerical terms, a clinically meaningful difference in blood pressure might be very different to a clinically meaningful difference in hemoglobin A1C. We first performed a paired-sample Wilcoxon test on the complete data set. For further metrics, we needed to scale the results to accurately compare outcomes which are measured on different scales, to avoid extreme values. This was achieved using the 95% CI/CrI of the point estimate that had been derived from the standard anchored ITC. The difference between the MAIC and ITC estimates was divided by the absolute value of the difference in the upper or lower limit of the CI/CrI and the ITC point estimate. We refer to this as the standardised difference in relative-efficacy estimates (SDIRE). For outcomes which used a ratio such as odds ratio or hazard ratio, the values were first transformed to the natural log scale.

In this analysis, the numerator has already accounted for the superiority definition in the treatment estimate. A positive value indicates that the MAIC relative-efficacy estimate for the IPD treatment was more favourable than the respective standard anchored ITC relative-efficacy estimate for the IPD treatment. Conversely, a negative value indicates that the MAIC relative-efficacy estimate for the AgD treatment was more favourable than the respective standard anchored ITC relative-efficacy estimate for the AgD treatment. Thus, the SDIRE is given by:

If an outcome is favourable (e.g. median overall survival time or rate of remission):

$$\frac{MAIC_{estimate} - ITC_{estimate}}{|ITC_{estimate} - ITC_{lower}|};\ where\ MAIC_{estimate} < ITC_{estimate}$$

$$\frac{MAIC_{estimate} - ITC_{estimate}}{|ITC_{upper} - ITC_{estimate}|};\ where\ MAIC_{estimate} > ITC_{estimate}$$

$$0; where\ MAIC_{estimate} = ITC_{estimate}$$

If an outcome is unfavourable (e.g. hazard ratio for progression or rate of stroke):

$$\frac{ITC_{estimate} - MAIC_{estimate}}{|ITC_{estimate} - ITC_{lower}|};\ where\ MAIC_{estimate} < ITC_{estimate}$$

$$\frac{ITC_{estimate} - MAIC_{estimate}}{|ITC_{upper} - ITC_{estimate}|};\ where\ MAIC_{estimate} > ITC_{estimate}$$

$$0; where\ MAIC_{estimate} = ITC_{estimate}$$

where $ITC_{estimate}$ denotes the point estimate derived from the standard anchored ITC, $MAIC_{estimate}$ denotes the point estimate derived from the MAIC, ITC_{lower} denotes the lower bound of the 95% CI/CrI for the ITC, and ITC_{upper} denotes the upper bound of the 95% CI/CrI for the ITC.

Results less than zero favour the AgD treatment, while results greater than zero favour the IPD treatment. Values outside the range $[-1,1]$ indicates that the MAIC estimate lies outside

the 95% CI/CrI for the ITC, which indicates that the adjustment has had quite a sizable impact on the original analysis. Because of the need to standardise comparisons on different scales using the 95% ITC CI/CrI, subrecords that did not report a 95% CI/CrI were excluded. A second analysis included the same formula above but scaled by the 95% CI of the MAIC estimate.

Results

We identified 168 subrecords where both the standard anchored ITC and MAIC estimates were available from 30 records [20-49]. The final PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram is shown in **Error! Reference source not found.**. Note that the PRISMA only provides details on the number of citations/records and does not provide a breakdown of subrecords. The results of these 168 subrecords are available in tabular format in Table S1 in the supplementary material. Note that the unique record numbering of the original systematic review has been maintained. Out of the total 168 subrecords, 109 subrecords favoured the treatment for which IPD was available, while 53 subrecords favoured the treatment with only AgD available. There were 6 subrecords where the estimate from the MAIC matched the estimate from the IPD. Using a paired-sample Wilcoxon test on all 168 records we found that there was a statistically significant difference at the 5% level in favour of the IPD treatment after the MAIC adjustment ($p<0.001$).

To report the SDIRE we included only studies where the CI/CrI of the ITC estimate was available. A second analysis used the CI of the MAIC estimate to include more trials. For the analysis using the CI/CrI of the ITC estimate there were 80 subrecords where both the standard anchored ITC and MAIC estimates were available, together with the ITC 95% CI/CrI. These 80 subrecords came from 17 records. These subrecords are shaded in dark grey in Table S1. Two records ([21] and [30]) fell into category 5 in our protocol in that they presented differing MAIC results of the same NMA. Due to the relatively small number of total records, it was decided to also include the earlier records in our final results.

For the analysis of standardising by the CI of the MAIC estimate an additional 20 subrecords were available from five records. These additional subrecords are shaded in light grey in Table S1. One of these records[41] contained two separate MAICs; one MAIC compared etanercept to adalimumab, while the other MAIC compared etanercept to certolizumab pegol.

Error! Reference source not found. shows the results and sensitivity analyses of our analysis when scaling by the CI/CrI of the ITC estimate. (**Error! Reference source not found.**A) shows all 80 subrecords. Subrecords are ordered in terms of most favourable differences reported following adjustment. If MAIC analyses were being reported in a systematic basis across the literature, so that favourable and unfavourable comparisons relative to the results of the standard anchored ITCs were equally likely to appear, then we

might reasonably expect this graph to have a broadly symmetric appearance around zero. This is not the case.

The median SDIRE of the subrecords (indicated by the maroon line) was 0.24. The mean SDIRE of the subrecords (indicated by the blue line) was 0.34. None of the 80 subrecords returned a result less than -1; indicating that no MAIC point estimate lay beyond the ITC 95% CI/CrI favouring an AgD treatment. There were six subrecords where the result was greater than 1; indicating that the MAIC point estimate for each of these comparisons was extreme enough to fall outside the ITC 95% CI/CrI, in favour of the IPD treatment.

In a secondary analysis we use a single representative subrecord for each publication. This provides a more balanced representation of MAIC use across all identified publications. To represent each record faithfully we used the median subrecord result (in terms of difference between ITC and MAIC efficacy estimate) from each record. Results of this analysis are shown in **Error! Reference source not found.**B. A similar pattern is clearly present. The median SDIRE of this analysis was 0.16. The mean SDIRE was 0.23.

Further sensitivity analyses, which used the least and most favourable subrecords for each publication, are shown in **Error! Reference source not found.**C and **Error! Reference source not found.**D. Median differences in these cases were -0.01 and 0.30 respectively, while the mean differences were 0.09 and 0.52 respectively.

The same analysis standardising by the MAIC CI is shown in **Error! Reference source not found.**. **Error! Reference source not found.**A shows all 100 subrecords. The median difference was 0.21 and the mean difference was 0.31. **Error! Reference source not found.**B uses the median (in terms of difference between ITC and MAIC efficacy estimate) subrecord result from each record. In this case the median difference was 0.16 and the mean difference was 0.18. **Error! Reference source not found.**C and **Error! Reference source not found.**D use the least and most favourable subrecords for each record. Median differences in these cases were 0.07 and 0.27 respectively, while the mean differences were -0.09 and 0.52 respectively. There were nine subrecords where the result was greater than 1; indicating that the ITC point estimate for each of these comparisons was extreme enough to fall outside the MAIC 95% CI, in favour of the IPD treatment. There was one subrecords where the result was less than -1; indicating that the ITC point estimate for each of these comparisons was extreme enough to fall outside the MAIC 95% CI, in favour of the AgD treatment.

The mean and median SDIRE from **Error! Reference source not found.** and **Error! Reference source not found.** are summarised in Table 2.

Discussion

Our results come from publications concerning a broad range of disease states and interventions until 25 February 2019. These results indicate a tendency to report MAIC results that are more favourable to the treatment for which IPD is available, relative to the results of the standard anchored ITCs, indicating a possible publication bias towards the IPD treatment. This may be caused by some combination of the publication of arguably inappropriate MAIC applications resulting in favourable outcomes, and the lack of publications when appropriate MAIC analyses did not lead to favourable outcomes. While it is not our goal to comment on the suitability of specific MAIC implementations, nor indeed on AgD comparisons where further adjustments were not made, this highlights that decision makers need to be careful when considering the suitability of the adjustments that are being made for the comparisons in question. We recommend continued observation of the usage of MAIC.

It is important to ensure that appropriate and correctly specified MAIC analyses are performed. Therefore, researchers, peer reviewers and HTA assessors must ensure that each MAIC is reported in a transparent manner, such that clear conclusions regarding any risks of bias can be drawn. Similarly, the analyst should question instances where no IPD adjustments have been made. The current status quo is that the absence of an MAIC is not routinely questioned, and this too can be problematic and may contribute to publication bias.

There are multiple limitations in our analysis, such as the reporting of multiple MAIC analyses in many of the publications, which couldyield conflicting results. We attempted to represent the information from each record faithfully by recording each outcome comparison as a subrecord. To investigate whether our conclusion could withstand considerable bias we undertook a sensitivity analysis where we used the subrecord result which had the most favourable result for the AgD treatment (i.e. the most conservative subrecord possible).

A second challenge was comparing data among a range of publications covering a diverse set of treatments and diseases, as well as the outcome types used. This is a challenge when reviewing a method as opposed to a treatment for a specific clinical indication, as it is a fundamentally different question. This has been attempted for questions such as the achieved power of a randomised controlled trial in the Cochrane database of systematic reviews[51], and for the heterogeneity of treatment effects within a meta-analysis [52]. We used a method of standardisation by the CI/CrI to assess results in an objective fashion. However, we recognise that there may be limitations in using one metric to represent multiple different outcomes, and limitations in capturing the nuances of individual treatment comparisons, such as the impact that the reported magnitude in difference could have in clinical practice. Furthermore, the SDIRE as currently implemented uses only the CI/CrI of either the MAIC or the ITC estimate, but not both at the same time. Further research could be undertaken to identify other metrics which may be more suitable and incorporate the uncertainty of both the MAIC and the ITC estimates.

This study did not evaluate the presence, magnitude, or direction of actual bias, i.e., relative to the true treatment effect. Furthermore, we did not evaluate the quality of conduct or of reporting of the underlying published MAIC/ITC studies. Therefore, it is not possible to say whether ITC or MAIC is closer to the truth in any given record or subrecord.

A limitation of this review is that it was performed in 2019. An updated review may improve generalizability.

Additionally, we note that many MAICs identified were unsuitable for our final appraisal. Many studies excluded were duplicates of later work, for example, conference presentations that formed the basis for journal publications. Still, there were a few studies which we excluded as they did not provide sufficient information. We cannot say for certain what effect the exclusion of these studies had on our results. However, it is likely that the excluded studies are of lower quality as their methods and results were not reported as clearly as the studies which have been included in our analysis. Nevertheless, the generalisability of our work is unknown, and any conclusions made here should be considered with caution.

We focused on MAIC only as this is routinely seen in published literature and HTA submissions. We explored using STCs publications also. However, only 13 subrecord results were available from three records. We note that other new methodology has also been developed in the area of population adjusted indirect comparisons [15,53]. We also note that MAIC and several other methods mentioned here account for the fact that only limited IPD is available, and will not be as useful in HTA as a full IPD NMA and network meta-regressions [54]. Similarly, if IPD is available for the two trials in question, utilising all the IPD by propensity scoring methods such as inverse probability treatment weighting is likely to yield more robust results.

Our review classes all standard anchored ITCs together. However, different types of ITC, for example, Bucher versus Bayesian, may differ in whether these produce less favourable results than MAIC.

Although our review focuses on MAICs compared to a standard anchored ITC, our method of analysis could be applied to other new and innovative methods, even outside the area of evidence synthesis. The timeframe of this systematic review reflects a period of uncertainty surrounding the benefits and drawbacks of MAIC, and provides a concrete example of how similar reviews for other innovative methods may be presented to HTA agencies in future.

This review highlights an important emerging issue in the reporting of MAIC analyses. We believe it might be useful for users of MAICs, to consider these findings in the case of anchored analysis, and carefully assess whether a standard anchored ITC, MAIC or alternative method of population adjustment is most appropriate for the available data.

We welcome the recent developments in the field of population adjustment, which has the potential to yield results specifically for the population under consideration. However, it is of the utmost importance that all methods and modelling choices are reported as transparently as possible, for a thorough assessment to be performed. There is a potential for greater work to be done to produce “best practice” procedures in terms of reporting, covariate selection, and

sensitivity analyses. We also recommend increased sharing of IPD to allow independent researchers to conduct more MAICs and other sensitivity analyses.

Conclusion

We have presented four analyses of SDIRE between the MAIC and the standard anchored ITC. We analysed all subrecords which fitted our research problem, and also examined the median subrecord, and the most and least favourable subrecord from each record. Most analyses show a positive mean and median result for the difference between the MAIC and standard anchored ITC estimate, indicating that, in general, published MAICs tend to be relatively more favourable to the IPD treatment. Additionally, we found six subrecords where the result was greater than 1 when stanardising by the CI/CrI of the ITC estimate; indicating that the MAIC point estimate for each of these comparisons was extreme enough to fall outside the standard anchored ITC 95% CI/CrI, in favour of the IPD treatment. No subrecord was found for the converse, that the MAIC point estimate was extreme enough to fall outside the ITC 95% CI/CrI in favour of the AgD treatment. When stanardising by the CI of the MAIC estimate we found nine subrecords where the result was greater than 1; and only one result was less than -1.

Although we recognise the importance of MAIC and other population adjustment methods, we recommend that results from these analyses are interpreted with caution. For anchored comparisons, we recommend:

1. The results of an MAIC are compared with the results of the standard anchored ITC to quantify the degree of adjustment made.
2. HTA analysts should demonstrate any potential biases in the standard anchored ITC (for example due to unbalanced effect modifiers in each arm) before performing an MAIC.
3. If no MAIC or other population-adjusted analysis is provided, decision-makers should carefully evaluate baseline characteristics across trials and the risks of bias in unadjusted ITCs.
4. Due to the subjective nature in terms of patient characteristic selection, we recommend that all MAIC analyses are reported as transparently as possible. A number of areas which often lack information when considering MAICs in HTA have been described previously. [50]

References

1. Caldwell DM, Ades A, Higgins J. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Bmj*. 2005;331(7521):897-900.
2. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*. 1997;50(6):683-691.
3. EUNetHTA. *Comparators & Comparisons: Direct and indirect comparisons*. 2013. *WP5*. February 2013.

4. Health Information and Quality Authority. *Guidelines for Evaluating the Clinical Effectiveness of Health Technologies in Ireland*. 2018.
5. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value in Health*. 2011;14(4):417-428.
6. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013*. 2013.
7. Lunny C, Veroniki A-A, Higgins JP, et al. Methodological review of NMA bias concepts provides groundwork for the development of a list of concepts for potential inclusion in a new risk of bias tool for network meta-analysis (RoB NMA Tool). 2022;
8. Signorovitch JE, Wu EQ, Andrew PY, et al. Comparative effectiveness without head-to-head trials. *PharmacoEconomics*. 2010;28(10):935-945.
9. Jansen JP, Trikalinos T, Cappelleri JC, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value in Health*. 2014;17(2):157-173.
10. Phillippo DM, Dias S, Ades A, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. *Statistics in medicine*. 2020;39(30):4885-4911.
11. Remiro-Azócar A, Heath A, Baio G. Conflating marginal and conditional treatment effects: Comments on "Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study". *Statistics in Medicine*. 2021;40(11):2753-2758.
12. Cheng D, Ayyagari R, Signorovitch J. The statistical performance of matching-adjusted indirect comparisons: Estimating treatment effects with aggregate external control data. 2020;
13. Cheng D, Tchetgen ET, Signorovitch J. On the Double-Robustness and Semiparametric Efficiency of Matching-Adjusted Indirect Comparisons. *Research Synthesis Methods*. 2022;
14. Remiro-Azócar A, Heath A, Baio G. Methods for population adjustment with limited access to individual patient data: A review and simulation study. *Research synthesis methods*. 2021;12(6):750-775.
15. Remiro-Azócar A, Heath A, Baio G. Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data. *Research Synthesis Methods*. 2022;13(6):716-744.
16. Leahy J, Walsh C. Assessing the impact of a matching-adjusted indirect comparison in a Bayesian network meta-analysis. *Research synthesis methods*. 2019;10(4):546-568.
17. Garritty C, Gartlehner G, Nussbaumer-Streit B, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *Journal of Clinical Epidemiology*. 2021/02/01/ 2021;130:13-22. doi:https://doi.org/10.1016/j.jclinepi.2020.10.007
18. Sutton A, Clowes M, Preston L, Booth A. Meeting the review family: exploring review types and associated information retrieval requirements. 2019;36(3):202-222. doi:https://doi.org/10.1111/hir.12276
19. Tricco AC, Langlois E, Straus SE. *Rapid reviews to strengthen health policy and systems: a practical guide*. 2017. 9241512768.
20. Bourdin A, Husereau D, Molinari N, et al. Matching-adjusted indirect comparison of benralizumab versus interleukin-5 inhibitors for the treatment of severe asthma: a systematic review. *European Respiratory Journal*. 2018;52(5)

21. Forsythe A. Matching-adjusted indirect treatment comparison of ribociclib and palbociclib as first-line treatments for HR1, HER2– ABC. *Annals of Oncology*. 2017;
22. Fox RJ, Chan A, Zhang A, et al. Comparative effectiveness using a matching-adjusted indirect comparison between delayed-release dimethyl fumarate and fingolimod for the treatment of multiple sclerosis. *Current medical research and opinion*. 2017;33(2):175-183.
23. Ishak KJ, Rael M, Hicks M, Mittal S, Eatock M, Valle JW. Relative effectiveness of sunitinib versus everolimus in advanced pancreatic neuroendocrine tumors: an updated matching-adjusted indirect comparison. *Journal of Comparative Effectiveness Research*. 2018;7(10):947-958.
24. Proskorovsky I, Su Y, Fahrbach K, et al. Indirect treatment comparison of inotuzumab ozogamicin versus blinatumomab for relapsed or refractory acute lymphoblastic leukemia. *Advances in therapy*. 2019;36(8):2147-2160.
25. Shafrin J, Shrestha A, Chandra A, Erder MH, Sikirica V. Evaluating Matching-Adjusted Indirect Comparisons in Practice: A Case Study of Patients with Attention-Deficit/Hyperactivity Disorder. *Health Economics*. 2017;26(11):1459-1466.
26. Signorovitch J, Swallow E, Kantor E, et al. Everolimus and sunitinib for advanced pancreatic neuroendocrine tumors: a matching-adjusted indirect comparison. *Experimental hematology & oncology*. 2013;2(1):1-8.
27. Signorovitch JE, Wu EQ, Betts KA, et al. Comparative efficacy of nilotinib and dasatinib in newly diagnosed chronic myeloid leukemia: a matching-adjusted indirect comparison of randomized trials. *Current medical research and opinion*. 2011;27(6):1263-1271.
28. Signorovitch JE, Wu EQ, Swallow E, Kantor E, Fan L, Gruenberger J-B. Comparative efficacy of vildagliptin and sitagliptin in Japanese patients with type 2 diabetes mellitus. *Clinical drug investigation*. 2011;31(9):665-674.
29. Song J, Ma Q, Gao W, et al. Matching-adjusted indirect comparison of blinatumomab vs. inotuzumab ozogamicin for adults with relapsed/refractory acute lymphoblastic leukemia. *Advances in Therapy*. 2019;36(4):950-961.
30. Tremblay G, Chandiwana D, Dolph M, Hearnden J, Forsythe A, Monaco M. Matching-adjusted indirect treatment comparison of ribociclib and palbociclib in HR+, HER2− advanced breast cancer. *Cancer management and research*. 2018;10:1319.
31. Tremblay G, Holbrook T, Milligan G, Pelletier CL, Rietschel P. Matching-adjusted indirect treatment comparison in patients with radioiodine-refractory differentiated thyroid cancer. *Comparative Effectiveness Research*. 2016;6:13-21.
32. Tremblay G, Westley T, Arondekar B, et al. Evaluating Different Indirect Treatment Comparison Approaches: A Case Study in Acute Myeloid Leukemia Patients Ineligible to Receive Intensive Chemotherapy. *Value in Health*. 2018;21:S17.
33. Van Sanden S, Baculea S, Diels J, Cote S. Comparative efficacy of ibrutinib versus obinutuzumab+ chlorambucil in first-line treatment of chronic lymphocytic leukemia: a matching-adjusted indirect comparison. *Advances in therapy*. 2017;34(7):1650-1661.
34. Warren R, Brnabic A, Saure D, et al. Matching-adjusted indirect comparison of efficacy in patients with moderate-to-severe plaque psoriasis treated with ixekizumab vs. secukinumab. *British Journal of Dermatology*. 2018;178(5):1064-1071.
35. Wu J, Saure D, Schacht A, et al. Matching-adjusted indirect comparisons of ixekizumab and secukinumab using etanercept and ustekinumab bridge comparators. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA; 2018:44-44.
36. Zigler JE, Blumenthal SL, Guyer RD, Ohnmeiss DD, Patel L. Progression of adjacent-level degeneration after lumbar total disc replacement: results of a post-hoc analysis

of patients with available radiographs from a prospective study with 5-year follow-up. *Spine*. 2018;43(20):1395.
37. Signorovitch JE, Betts KA, Reichmann WM, et al. One-year and long-term molecular response to nilotinib and dasatinib for newly diagnosed chronic myeloid leukemia: a matching-adjusted indirect comparison. *Current medical research and opinion*. 2015;31(2):315-322.
38. Ishak K, Phatak H, Rael M, et al. A simulated head to head comparison of stroke and major bleeding with apixaban versus rivaroxaban in high-risk NVAF patients. OXFORD UNIV PRESS GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND; 2015:743-744.
39. Sikirica V, Findling RL, Signorovitch J, et al. Comparative efficacy of guanfacine extended release versus atomoxetine for the treatment of attention-deficit/hyperactivity disorder in children and adolescents: applying matching-adjusted indirect comparison methodology. *CNS drugs*. 2013;27:943-953.
40. Betts K, Mittal M, Song J. Relative efficacy of adalimumab versus secukinumab in active ankylosing spondylitis: a matching-adjusted indirect comparison. Abstract OP115 presented at the 25th European League Against Rheumatism Congress. 2016, June 8-11. 2016:
41. Benedict A, Ishak J, Gal P, et al. THU0207 Comparative Analysis of the Effectiveness of anti-TNF Therapies in Non-Radiographic Axial Spondyloarthritis Using Novel Statistical Techniques. BMJ Publishing Group Ltd; 2015.
42. 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;28:935-945.
43. Chan A, Cutter G, Fox RJ, Xiao J, Lewin JB, Edwards MR. Comparative effectiveness of delayed-release dimethyl fumarate versus glatiramer acetate in multiple sclerosis patients: results of a matching-adjusted indirect comparison. *Journal of comparative effectiveness research*. 2017;6(4):313-323.
44. Signorovitch J, Erder MH, Xie J, et al. Comparative effectiveness research using matching-adjusted indirect comparison: an application to treatment with guanfacine extended release or atomoxetine in children with attention-deficit/hyperactivity disorder and comorbid oppositional defiant disorder. *pharmacoepidemiology and drug safety*. 2012;21:130-137.
45. Strand V, Betts KA, Mittal M, Song J, Skup M, Joshi A. Comparative effectiveness of adalimumab versus secukinumab for the treatment of psoriatic arthritis: a matching-adjusted indirect comparison. *Rheumatology and therapy*. 2017;4:349-362.
46. Xie J, Juday T, Swallow E, et al. IN2 Comparative Efficacy at 48 Weeks of Atazanavir/Ritonavir Versus Darunavir/Ritonavir in Treatment Naive Hiv-1 Patients: A Matching Adjusted Indirect Comparison of Randomized Trials. *Value in Health*. 2012;15(4):A10.
47. Xie J, Juday T, Swallow E, et al. Matching-adjusted indirect comparison of lipid profile at 48 weeks among treatment naïve HIV-1 patients treated with Atazanavir/Ritonavir versus Darunavir/Ritonavir. *Value in Health*. 2013;16(3):A79.
48. Fautrel B, Zhu B, Taylor PC, et al. SAT0225 Comparative effectiveness in pain and haq-di improvement for baricitinib versus adalimumab, tocilizumab, and tofacitinib monotherapies in csdmard-naÏve rheumatoid arthritis patients: a matching-adjusted indirect comparison (MAIC). BMJ Publishing Group Ltd; 2018.
49. Burkhardt N, Strober B, Brnabic A, et al. Indirect comparison of ixekizumab (IXE) and secukinumab (SEC) using matched-adjusted indirect comparisons (MAIC). WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA; 2017:43-43.

50. Holmes EM, Leahy J, Walsh CD, White A, Donnan PT, Lamrock F. Difficulties arising in reimbursement recommendations on new medicines due to inadequate reporting of population adjustment indirect comparison methods. *Research Synthesis Methods*. 2019;10(4):615-617.
51. van Zwet E, Schwab S, Senn S. The statistical properties of RCTs and a proposal for shrinkage. *Statistics in Medicine*. 2021;40(27):6107-6117.
52. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International journal of epidemiology*. 2012;41(3):818-827.
53. Phillippo DM, Dias S, Ades A, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2020;183(3):1189-1210.
54. Riley RD, Dias S, Donegan S, et al. Using individual participant data to improve network meta-analysis projects. *BMJ Evidence-Based Medicine*. 2022;

Table 1: Protocol for record selection

1	The set of outcomes reported in each record was identified.
2	Records which report fewer outcomes that have already been published in an alternative record were excluded in favour of the more comprehensive publication. However, duplicate records were checked for data consistency and where disparities exist, these were reported and considered for sensitivity analysis, as mentioned in point 5.
3	Only in the case where there is an additional outcome not reported elsewhere was more than one record per treatment comparison included in the review.
4	Where multiple records for the same treatment comparison exist and neither record contained any additional information with respect to outcomes, one record was selected based on one of the following distinguishing criteria (in order): •Publication date (latest preferred) •Type of publication (in order of preference): i. Journal publication ii. Conference poster or HTA agency / EMA report iii. Conference/published abstract
5	Where a disparity in published results exists, later publications took precedence under the assumption that results are more accurate. However, differing results from earlier publications were considered for sensitivity analysis within the review.

Table 2: Mean and median SDIRE standardising by either the ITC or MAIC CI/CrI

		Standardising by the ITC CI/CrI	Standardising by the MAIC CI
Median SDIRE	All subrecords	0.24	0.21
	Median subrecord	0.16	0.16
	Lowest subrecord	-0.01	0.07
	Highest subrecord	0.3	0.27
Mean SDIRE	All subrecords	0.34	0.31
	Median subrecord	0.23	0.18
	Lowest subrecord	0.09	-0.09
	Highest subrecord	0.52	0.52