www.ispor.org

ISPOR
Improving healthcare decisions

ISPOR Special Interest Group

Workshop 21: INDIRECT TREATMENT COMPARISONS: AN INTERACTIVE WORKSHOP ON CHOOSING THE RIGHT TOOL FOR THE AVAILABLE DATA

ISPOR Europe 2018 | Barcelona, Spain

Wednesday, 14 November 2018 | 15:00 – 16:00

ISPOR Statistical Methods in HEOR Special Interest Group (SIG)

- **Mission:** To provide statistical leadership for strengthening the use of appropriate statistical methodology in health economics and outcomes research and improve the analytic techniques used in real world data analysis.

Co-Chairs of SIG

- **Rita M. Kristy, MS**, Senior Director, Medical Affairs Statistics, Astellas Pharma Global Development, Northbrook, IL, USA
- **David J. Vanness, PhD,** Professor, Health Policy and Administration, The Pennsylvania State University, State College, PA, USA

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Rate your experience with ITCs:

3

ISPOR www.ispor.org

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Which terms are you NOT familiar with?

4

Rita M. Kristy, MS
Senior Director, Medical Affairs Statistics, Astellas Pharma Global Development

Introduction – Indirect Treatment Comparisons (ITC)

ISPOR

www.ispor.org

Introduction – Indirect Treatment Comparisons (ITC)

- Systematic reviews of randomized controlled trials (RCTs) are a standard method of analyzing information in the health-care setting.
  - ITCs are often necessary in order to combine this information and answer many research questions of interest.
    - This is particularly important in the comparative effectiveness landscape where head-to-head comparisons of interest are often unavailable.
- Approach:
  - ITCs often use the relative effects of the treatments versus their common comparator (e.g., placebo) in order to assess the head-to-head comparison of interest

Assumptions

- Because ITCs require the relative effect of treatments with their common comparators:
  - This assumes that the common comparators (e.g., control groups) are sufficiently similar to make the combination of relative effects viable.
  - The studies used for the ITC are sufficiently similar.
    - e.g., There would be challenges combining a pediatric-only treatment with an exclusively adult treatment.
    - The relative effects of the treatments to their common comparator may be influenced/biased/imbalanced differently by their different patient populations.

7

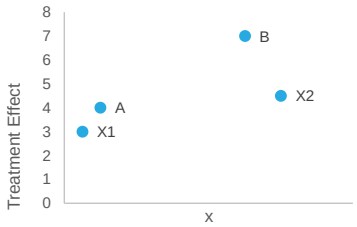
Assumptions cont.

- Concern exists about bias resulting from misuse.
  - e.g, Heterogeneity: patient populations with different comorbidity burdens, at different points in the disease state, different study conduct, etc.
  - It is possible for the studies to be different in unobservable ways that cannot be adjusted for.
- Randomization:
  - The properties of randomization hold within the individual studies
  - This does not extend across studies. This means that studies may differ more than white noise in characteristics such as patient demographics.
  - If there are imbalanced characteristics that are related to the treatment effect, then the studies are known as heterogeneous. [1]
- ITCs may be unbiased if assumptions are met – homogeneity, similarity of studies, consistency of evidence

8

Assumptions cont.

- ITCs should not "break randomization" [1, 2]
  - Bucher et al. (1997) studied this and proposed that treatment comparisons be based on the relative treatment effects from each study and not the raw/direct results
  - Example: Compare treatments A and B
  - Study 1 compared A with placebo, X (X1), and Study 2 compared B with X (X2)



  - A and B should be compared only through the relative A vs X1 and B vs X2 differences. This takes into account the placebo effects from the different studies.

10

Fixed and Random Effects models [1]: Estimating tr_{AB}

- Fixed-Effects model
  - It is assumed that there is no variation in relative treatment effects across studies for a particular pairwise comparison [3, 4]. Observed differences for a particular comparison among study results are solely due to chance.
- If there is heterogeneity across studies, however, a Fixed-Effects model should not be used.
- Random-Effects model
  - Assumes that treatment effects across studies are considered *exchangeable* (i.e., the prior position of expecting underlying effects to be similar but not identical) and can be described as a sample from a distribution whose mean is the pooled relative effect and whose SD reflects the heterogeneity [5-9].

11

Common Methods - ITCs

- In 2011, an ISPOR Task Force on ITCs published some guidelines on analyses and noted that although Random-Effects models account for heterogeneity, it does not explain the source of the heterogeneity in the data. [1]
- Similarly, extending a fixed- or random-effects model by incorporating treatment-by-covariate interaction terms can also improve model fit and explanations of heterogeneity.
- Direct and indirect comparisons can be combined to improve statistical power.

12

MAIC vs. Bucher

- Bucher's Adjusted Indirect Comparisons method [2]
  - Assumes the relative effect of a treatment is the same across studies
  - Uses relative effects to compare treatments (advocated by the 2011 ISPOR Task Force) [1]
  - Does not address treatment effect modifiers (treatment confounders) that may be different between studies
- Matching-Adjusted Indirect Comparison (MAIC) method
  - The purpose is to balance observable treatment confounders between studies
  - Requires individual patient data (IPD) for one of the treatment arms of interest
    - This is not uncommon in HTA submissions with IPD available for the sponsor studies but not for the comparator studies
  - Propensity-score weights are used such that the IPD group is similar with the aggregated patient characteristics for the comparator arm study(s)

13

References

1. Jansen, Jeroen P., 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 14.4 (2011): 417-428.
2. Bucher, Heiner C., et al. "The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials." Journal of clinical epidemiology 50.6 (1997): 683-691.
3. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health 2008;11:956–64.
4. Borenstein M, Hedges LV, Higgins JPT, Rothstein H. Introduction to Meta-Analysis. Chichester, England: John Wiley & Sons, Ltd., 2009.
5. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from: http://www.cochranehandbook. org. [Accessed February 11, 2010].
6. Borenstein M, Hedges LV, Higgins JPT, Rothstein H. Introduction to Meta-Analysis. Chichester, England: John Wiley & Sons, Ltd., 2009.
7. Skene AM, Wakefield JC. Hierarchical models for multicentre binary response studies. Stat Med 1990;9:919–29.
8. Gelman AB, Carlin JS, Stern HS, Rubin DB. Bayesian Data Analysis. Boca Raton, FL: Chapman and Hall–CRC; 1995.
9. Cappelleri JC, Ioannidis JPA, Lau, J. Meta-analysis of therapeutic trials. In: Chow S-C, ed., Encyclopedia of Biopharmaceutical Statistics (3rd ed.), Revised and Expanded. New York, NY: Informa Healthcare, 2010.

Necdet Gunsoy, PhD, MPH
Director, Analytics and Innovation GlaxoSmithKline (GSK)

www.ispor.org

Anchored Indirect Treatment Comparison

Standard method or MAIC? A case study

Disclosure

I am a GSK employee and hold GSK stock. The case study presented is based upon data for a marketed GSK product.

The content of this presentation is a reflection of my personal views and do not necessarily reflect the views of GSK.

Any details relating to this case study (treatment names, outcomes, results) have been modified for the purposes of this presentation. Thus, the case study is entirely fictive.

16

Pre-requisites

Methods:

- Bucher method
- MAIC

Terminology:

- Anchored network
- Prognostic factor
- Effect modifier

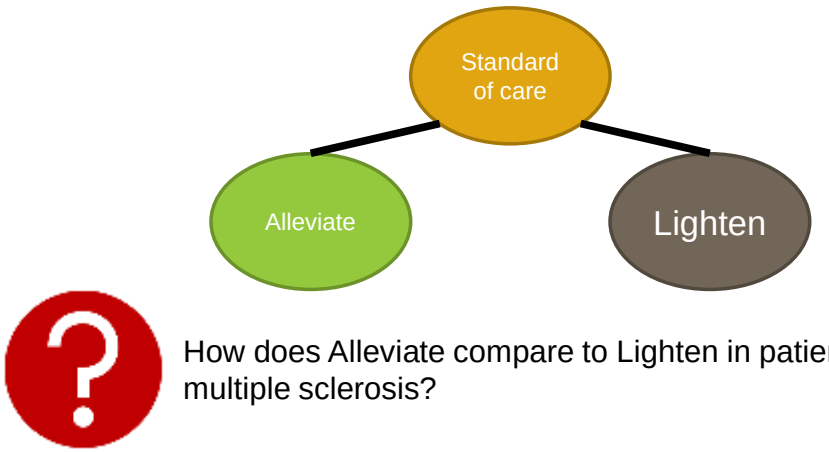
17

Context

- Two treatments (Alleviate and Lighten) are licensed for patients with multiple sclerosis (MS)
- Alleviate and Lighten have a similar method of action
- There is at least one effect modifier which impacts the expected treatment effect
- The patients included in the RCTs for the two treatments are different in terms of the distribution of effect modifier(s)
- Both treatments were compared against the same standard of care (i.e. anchored indirect treatment comparison)

18

Context



How does Alleviate compare to Lighten in patients with multiple sclerosis?

19

Your job (as the analyst)

Can the two interventions be compared?

What method is most appropriate?

Would the result be different either way?

20

Can the two interventions be compared?

Poll question 1

What should we evaluate first?

- What data are available?
- What are the effect modifier(s)?

21

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: What should be evaluated first?

22

Can the two interventions be compared?

- What are the effect modifier(s)?

- Both interventions have a similar method of action (MOA)
- There is one confirmed effect modifier related to the MOA
- There is mixed evidence relating to further effect modifier(s)

23

Can the two interventions be compared?

- What are the effect modifier(s)?

Poll question 2

Is this characteristic an effect modifier?

Outcome: Reduction in relapse events (Rate ratio)

Levels of biomarker	Alleviate	Lighten
Low	0.75	0.75
Medium	0.50	0.60
High	0.30	0.45

24

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Is this characteristic an effect modifier?

25

Can the two interventions be compared?

- What are the effect modifier(s)?

Concept: "Shared effect modification" assumption

When two treatments have similar MOA, it would be expected that effect modifier(s) are shared. If true, this would imply that

- All effect modifiers are the same for the two treatment
- All effect modifiers impact treatment effects in the same way for both treatment

Levels of biomarker	Alleviate	Lighten
Low	0.75	0.75
Medium	0.50	0.60
High	0.30	0.45

Phillippo et al, NICE DSU TSD18 (2016)

Can the two interventions be compared?

- What are the effect modifier(s)?

Poll question 3

Is this characteristic an effect modifier?

Outcome: Reduction in relapse events (Rate ratio)

Presence of comorbidity	Alleviate	Lighten
Yes	0.70	0.35
No	0.50	0.65

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Is this characteristic an effect modifier?

28

Can the two interventions be compared?

- What are the effect modifier(s)?

Conflicting evidence for effect modification?

This characteristic may be an effect modifier, but results from the two interventions appear to conflict.

- Seek clinical advice
- Assess the implications
- How likely is it that the characteristic is an effect modifier?

Presence of comorbidity	Alleviate	Lighten
Yes	0.75	0.35
No	0.50	0.65

29

Can the two interventions be compared?

- What are the effect modifier(s)?

Implications of conflicting evidence for effect modification

What happens if an effect modifier with conflicting evidence is matched for?

Presence of comorbidity	Alleviate		Lighten	
	%	RR	%	RR
Yes	30	0.70	15	0.35
No	70	0.50	85	0.65
Overall	100	0.56	100	0.61

Conclusions differ depending on who does the matching

0.56 — Matched to Alleviate — 0.56

0.53 — Matched to Lighten — 0.61

30

Can the two interventions be compared?

- What are the effect modifier(s)?

- There is one confirmed effect modifier
- There is one potential effect modifier, but results by the effect modifier conflict between the two interventions

The existence of conflicting effect modification can have **serious** implications on the interpretability of results:

- Shared effect modification assumption violated
- Difference in results depending on which direction the matching is performed

Conclusions differ depending on who does the matching

Lighten 0.56 — Matched to Alleviate — 0.56 **Lighten**

Alleviate 0.53 — Matched to Lighten — 0.61 **Alleviate**

31

Can the two interventions be compared?

- What data are available?

- Intent-to-treat (ITT, i.e. overall population)
  - The patients differ in terms of effect modifier(s)
  - Baseline characteristics for these effect modifier(s) are reported

- Subgroup
  - Subgroup results are available for effect modifier(s), but individually
  - The patient *may* still differ in terms of other effect modifier(s)
  - Baseline characteristics in these subgroups are not available

32

ISPOR www.ispor.org

Can the two interventions be compared?

- What data are available?

ITT	Alleviate	Lighten
	0.56	0.61

Levels of biomarker	Alleviate		Lighten	
	%	RR	%	RR
Low	40	0.75	45	0.75
Medium	40	0.50	15	0.60
High	20	0.30	40	0.45

Same direction
Different magnitude

Presence of comorbidity	Alleviate		Lighten	
	%	RR	%	RR
Yes	30	0.70	15	0.35
No	70	0.50	85	0.65

Different direction
Different magnitude

33

Can the two interventions be compared?

- What data are available?

We can

- Perform a standard (“unadjusted”) ITC on the ITT population
- Perform a standard (“unadjusted”) ITC on subgroups
- Perform a matched (MAIC) comparison on the ITT population

We cannot

- Perform a matched (MAIC) comparisons on subgroups

34

What method is most appropriate?

Poll question 4

What method is most appropriate?

- Bucher method on ITT
- Bucher method on subgroup data
- MAIC in the ITT population

35

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: What method is most appropriate?

36

What method is most appropriate?

- Bucher method on ITT

"Unadjusted" comparison on ITT

- We know that patient populations differ in terms of effect modifier(s) distribution across studies
- We know that an 'unadjusted' comparison on ITT populations would be biased by differences in baseline characteristics

37

What method is most appropriate?

- Bucher method on subgroups

"Unadjusted" comparison in subgroups

- We have outcome data within subgroups by effect modifier(s) separately
- We know that results broken down by effect modifier(s) may be conflicting and/or have different magnitudes
- We don't know if the distribution of any other potential effect modifier(s) are balanced within subgroups

38

What method is most appropriate?

- MAIC in the ITT population

Adjusted (MAIC) comparisons in the ITT population

- We can adjust for differences in the distribution of effect modifier(s)
- We know that matching one way or the other might lead to different results

39

What method is most appropriate?

	Bucher ITT	Bucher Subgroups	MAIC ITT
Are the populations being compared the same?	✗	✗ ✓	✓
Are the results generalisable to the overall population?	✗	✗ ✓	✗ ✓
Are the results relevant for sub-populations?	✗	✓	✗

40

What method is most appropriate?

	Bucher ITT
Are the populations being compared the same?	✗
Are the results generalisable to the overall population?	✗
Are the results relevant for sub-populations?	✗

Comparisons between ITT populations will be biased as the populations are different in terms of effect modifier(s) distributions.

Results will lack credibility and validity, and are therefore not generalisable.

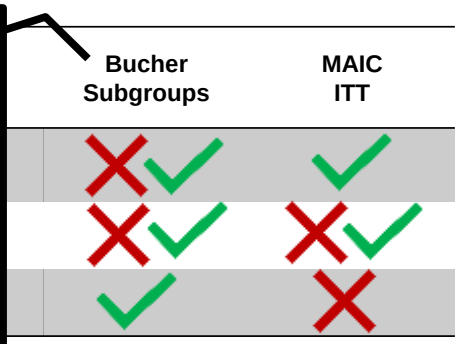
41

What method is most appropriate?

Comparisons between subgroups are balanced for an effect modifier, but potentially not for others, if others exist.

Results could be extrapolated to an overall population based upon a distribution of effect modifiers.

The validity and credibility of results may depend on the existence of other effect modifiers and whether or not their distributions differ within the subgroups.



42

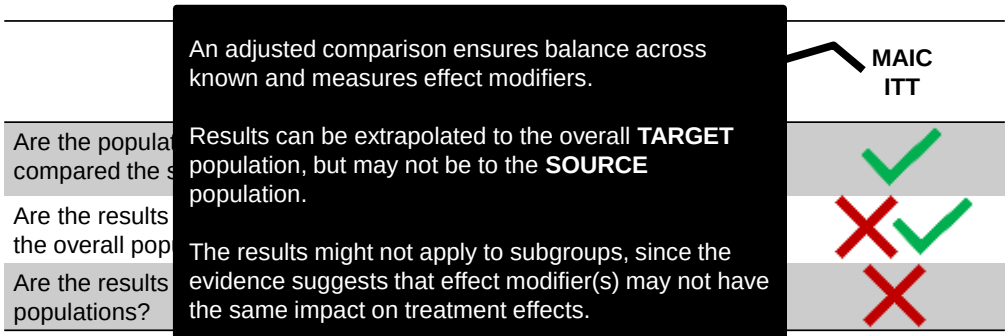
What method is most appropriate?

An adjusted comparison ensures balance across known and measures effect modifiers.

Results can be extrapolated to the overall **TARGET** population, but may not be to the **SOURCE** population.

The results might not apply to subgroups, since the evidence suggests that effect modifier(s) may not have the same impact on treatment effects.

The validity and credibility of results may depend on the selection of effect modifiers, and whether the shared effect modifier assumption is met.



43

What method is most appropriate?

Poll question 4 (Revisited)

What method is most appropriate?

- Bucher method on ITT
- Bucher method on subgroup data
- MAIC in the ITT population

Both methods are “equally” appropriate, as each have strengths and limitations associated with them.

44

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: What method is most appropriate (now)?

45

Would the results be different either way?

Poll question 5

If you did not have access to individual data, which method would you use?

- Bucher method on ITT
- Bucher method on subgroup data

ITT	Alleviate		Lighten	
	0.56		0.61	
Levels of biomarker	**Alleviate**		**Lighten**	
	%	RR	%	RR
Low	40	0.75	45	0.75
Medium	40	0.50	15	0.60
High	20	0.30	40	0.45
Presence of comorbidity	**Alleviate**		**Lighten**	
	%	RR	%	RR
Yes	30	0.70	15	0.35
No	70	0.50	85	0.65

46

ISPOR www.ispor.org

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: If you did not have access to individual data, which method would you use?

47

Would the results be different either way?

ITT	Alleviate		Lighten		ITC estimate (A vs. L)
	0.56		0.61		0.93

Alleviate reduces undesirable events by 7% versus Lighten

Level of biomarker	Alleviate		Lighten		ITC estimate (A vs. L)
	%	RR	%	RR	
Low	40	0.75	45	0.75	1.00
Medium	40	0.50	15	0.60	0.83
High	20	0.30	40	0.45	0.67

Presence of comorbidity	Alleviate		Lighten		ITC estimate (A vs. L)
	%	RR	%	RR	
Yes	30	0.70	15	0.35	2.00
No	70	0.50	85	0.65	0.77

Results varied dependent on the subgroup considered.

48

Would the results be different either way?

Poll question 6

If you were running an analysis on behalf of Alleviate, which method would you use?

- Bucher method on subgroup data
- MAIC using Alleviate individual data in the ITT population

ITT	Alleviate		Lighten	
	0.56		0.61	

Levels of biomarker	Alleviate		Lighten	
	%	RR	%	RR
Low	40	0.75	45	0.75
Medium	40	0.50	15	0.60
High	20	0.30	40	0.45

Presence of comorbidity	Alleviate		Lighten	
	%	RR	%	RR
Yes	30	0.70	15	0.35
No	70	0.50	85	0.65

49

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: If you were running an analysis on behalf of Alleviate, which method would you use?

50

Would the results be different either way?

Level of biomarker	Alleviate		Lighten		ITC estimate (A vs. L)
	%	RR	%	RR	
Low	40	0.75	45	0.75	1.00
Medium	40	0.50	15	0.60	0.83
High	20	0.30	40	0.45	0.67

There is inconclusive evidence to suggest that presence of comorbidity is an effect modifier.

Alleviate was associated with larger reductions in relapse events with increasing level of biomarker.

MAIC in ITT	Alleviate	Lighten	ITC estimate (A vs. L)
Unadjusted	0.56	0.61	0.93
Adjusted			
>Biomarker	0.53	0.61	0.87
> + comorbidity	0.53	0.61	0.86

With adjustment, Alleviate was shown to reduce the rate of relapse events by 14% versus Lighten versus 7% when unadjusted. Results did not change even if "presence of comorbidity" was also adjusted for.

51

Would the results be different either way?

Poll question 7

If you were running an analysis on behalf Of Lighten, which method would you use?

- Bucher method on subgroup data
- MAIC in the ITT population

ITT	Alleviate		Lighten	
	0.56		0.61	
Levels of biomarker	**Alleviate**		**Lighten**	
	%	RR	%	RR
Low	40	0.75	45	0.75
Medium	40	0.50	15	0.60
High	20	0.30	40	0.45
Presence of comorbidity	**Alleviate**		**Lighten**	
	%	RR	%	RR
Yes	30	0.70	15	0.35
No	70	0.50	85	0.65

52

ISPOR www.ispor.org

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: If you were running an analysis on behalf of Lighten, which method would you use?

53

Would the results be different either way?

Level of biomarker	Lighten	Alleviate	ITC estimate (L vs. A)
Low	0.75	0.75	1.00
Medium	0.60	0.50	1.20
High	0.45	0.30	1.50

Presence of comorbidity	Lighten	Alleviate	ITC estimate (L vs. A)
Yes	0.35	0.70	0.50
No	0.65	0.50	1.30

Lighten reduced the rate of relapse events versus Alleviate in patients with a presence of comorbidity.

MAIC in ITT	Lighten	Alleviate	ITC estimate (L vs. A)
Unadjusted	0.61	0.56	1.08
Adjusted	0.57	0.56	1.01

With adjustment, Lighten was no different from Alleviate

54

Concluding remarks

- Selecting a method is not as straightforward as it seems
- A number of methodological approaches may be appropriate for a particular analysis
- There is no guarantee two approaches will lead to the same conclusions
- There needs to be careful consideration on the interpretation of results based on the approach taken

55

Anthony Hatswell, MSc
Director
Delta Hat Limited

Matching Adjusted Indirect Comparison (MAIC): A simulation study to see if it works

Matching Adjusted Indirect Comparison (MAIC)

- Matching Adjusted Indirect Comparison was first proposed by Signorovitch et al. in 2010, and has been discussed in a NICE DSU report (TSD 18, Phillipo et al)
- The method has been discussed in depth in other presentations (and the recent literature).
- Fundamentally MAIC is propensity weighting to aggregate characteristics
  - This is useful where we do not have patient level data to make comparisons e.g. comparator trials, or studies identified in the literature
- Because the method is relatively new, it hasn't been used much, and we don't really know how it performs – both under ideal conditions, and when assumptions (either explicit or implicit) are violated
- My interest is primarily in uncontrolled studies (sometimes termed 'single arm trials') - it is therefore the unanchored form investigated here

Assumptions underpinning MAIC

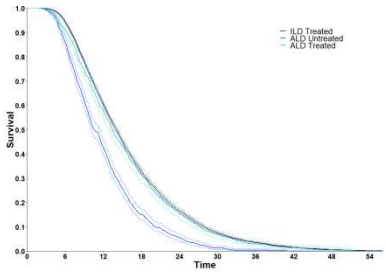
- In the method, patients from the study you do have (Individual Level Data, ILD) are reweighted such that their weighted characteristics match the characteristics for the study you don't have (Aggregate Level Data, ALD)
- When we reweight, what are we effectively assuming?
  - The patients are similar, just their characteristics appear in different proportions
  - There is considerable overlap between populations
  - The 'things' we are weighting for matter
  - The 'things' we are weighting for act in a linear fashion
  - We have enough patients to perform the reweighting
- What happens if you flex or violate these assumptions?

Design of the simulation study

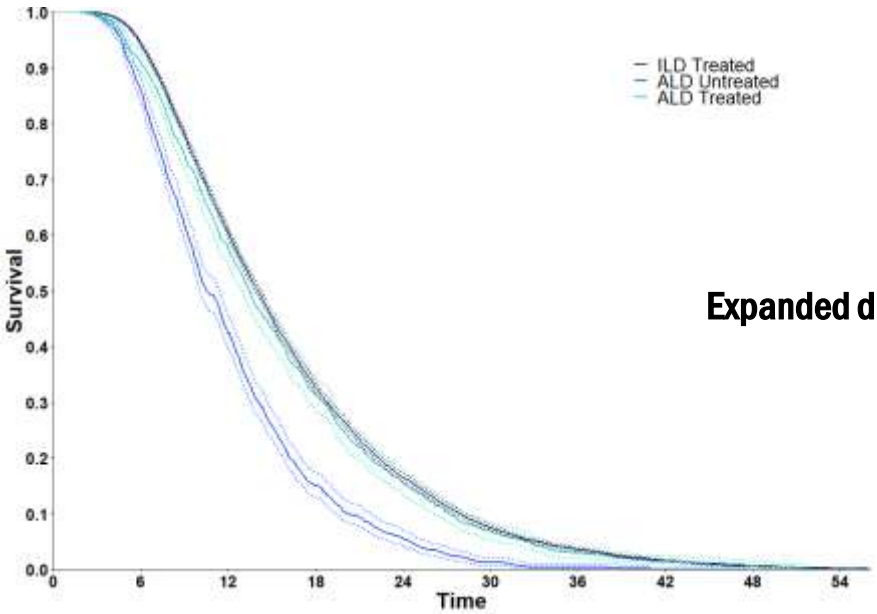
- We want to know if MAIC works as intended (and how well it works) under ideal conditions
- To do this we simulated studies with outcomes similar to those seen in advanced cancer (where MAIC has been mostly used)
  - Survival as an endpoint (an outcome of mean survival is used)
  - 1,000 patients in the ALD (which we only allow the method to see aggregate characteristics for)
  - 10,000 patients in the ILD (which we have access to), outcomes using a hazard ratio from the ALD
  - Matching on 4 covariates
  - Covariates slightly more favourable in the ILD versus the ALD:
    - ILD: $\sim N(mean\ 0.6,\ SD\ 0.15)$
    - ALD: $\sim N(mean\ 0.5, SD\ 0.15)$
- Runtime considerations:
  - 1,000 simulations takes about an hour on a very fast computer (Intel i5 8600K), which isn't unreasonable for an individual run
  - Adding lots of scenarios (with more runs) becomes a burden, but an individual run is not a problem

Simulation setup

- We have three lines on the graph
  - The blue line represents the outcomes for the ALD patients who were not treated
  - The black line represents the outcomes for the ILD patients who did receive treatment
  - The interesting one is the green line, which are the outcomes for what *would* have happened if the ALD patients had been treated (after all they have unfavourable characteristics)
    - Because this is a simulation study – we are able to calculate the counterfactual for these patients!



- The question is whether MAIC can reweight the black line, to be similar to the green one



Expanded diagram

Results of the basic simulation study

- Scenario setup:

	ALD Not Treated	ALD Treated	ILD Treated
Months	12.0	15.3	16.1

- Outcomes of interest:

Failed to converge	Bias (mean)	Bias (absolute)	Coverage	Worse than naive	Median bias % reduction

- The results are presented alongside a "naïve comparison" i.e. comparing the outcomes without performing any adjustment
  - No other methods are presented in the work as it stands

Results of the basic simulation study

- Results of the basic comparison, 1000 simulations

	Failed to converge	Bias (mean)	Bias (absolute)	Coverage	Worse than naive	Median bias % reduction
Naïve comparison		0.8	0.8	0	0	0

- "60% of the time, it works all the time" ('Brian Fantana', Anchorman, 2004)
  - One issue is we don't know when MAIC will and won't exaggerate bias
  - Fundamentally MAIC improves results more often than it worsens them, and appears to be unbiased
- I now have 3 slides of results – these are presented for completeness, and I will highlight the salient information

64

ISPOR www.ispor.org

	Failed to converge	Bias (mean)	Bias (absolute)	Coverage	Worse than naive	Median bias % reduction	
Exponential distribution used as the survival function							
Naïve comparison		0.82	0.82	0	0	0	
MAIC	0.00%	0	0.27	1	0.1	0.74	
Lognormal distribution used as the survival function							
Naïve comparison		0.8	0.8	0	0	0	
MAIC	0.00%	0	0.18	1	0.02	0.82	
All patient characteristics are binary variables							
Naïve comparison		0.82	0.82	0	0	0	
MAIC	0.00%	0.01	0.18	1	0.04	0.83	
Patient characteristics have low explanatory power							
Naïve comparison		0.23	0.3	0.18	0	0	
MAIC	0.00%	0.03	0.27	1	0.41	0.18	
Patient characteristics have high explanatory power							
Naïve comparison		1.62	1.62	0	0	0	
MAIC	0.00%	0	0.11	1	0	0.94	
Treatment effect is small (HR = 0.9)							
Naïve comparison		0.81	0.81	0	0	0	
MAIC	0.00%	0	0.18	1	0.03	0.82	
Treatment effect is huge (HR = 0.2)							
Naïve comparison		0.81	0.82	0.01	0	0	
MAIC	0.00%	-0.02	0.45	1	0.21	0.54	

65

ISPOR www.ispor.org

	Failed to converge	Bias (mean)	Bias (absolute)	Coverage	Worse than naive	Median bias % reduction	
Half matched parameters are nuisance parameters							
Naïve comparison		0.61	0.61	0	0	0	
MAIC	0.00%	0	0.19	1	0.09	0.75	
All matched parameters are nuisance parameters							
Naïve comparison		-0.02	0.17	0.89	0	0	
MAIC	0.00%	-0.02	0.17	1	0.47	0.01	
Parameters have a squared effect							
Naïve comparison		0.8	0.8	0	0	0	
MAIC	0.00%	0	0.19	1	0.03	0.8	
Parameters have a multiplicative effect							
Naïve comparison		0.74	0.74	0	0	0	
MAIC	0.00%	-0.01	0.28	1	0.13	0.69	
Covariate sampling is very close							
Naïve comparison		0.4	0.4	0	0	0	
MAIC	0.00%	-0.01	0.17	1	0.17	0.68	
Covariate sampling is not close							
Naïve comparison		2.42	2.42	0	0	0	
MAIC	0.00%	-0.02	1.54	1	0.2	0.48	
All parameters are correlated							
Naïve comparison		0.8	0.8	0	0	0	
MAIC	0.00%	0	0.18	1	0.03	0.82	

	Failed to converge	Bias (mean)	Bias (absolute)	Coverage	Worse than naive	Median bias % reduction
4 parameters matched, 2 unmatched (hidden) with the same distributions						
Naïve comparison		0.98	0.98	0	0	0
MAIC	0.00%	0.34	0.36	1	0	0.64
As above, but the parameters are correlated with the observed characteristics						
Naïve comparison		0.61	0.61	0	0	0
MAIC	0.00%	0	0.2	1	0.1	0.72
Trimmed characteristics in the ILD (no poor performers)						
Naïve comparison		1.27	1.27	0	0	0
MAIC	0.00%	0.04	0.53	1	0.06	0.63
Trimmed characteristics in the ALD (no good performers)						
Naïve comparison		1.27	1.27	0	0	0
MAIC	0.00%	0	0.28	1	0	0.81

As a summary

- MAIC performs reasonably well, and appears to be a substantial improvement on a naïve comparison, provided the main assumptions are met
  - Note: Other methods are available – do consider these also
- Be cautious if
  - Patient characteristics are not hugely important, or are missing
  - If the patient characteristics are not independent, or may interact in some way
  - You are missing information on a characteristic, but suspect it is correlated with something you do observe
- Be very cautious if
  - Non matching studies, such as different inclusion/exclusion criteria
  - It is not possible to compare studies on something important, and there is no surrogate available

Next steps

- MAIC is effectively a form of propensity weighting – I would like to include a comparison directly with propensity weighting
  - Propensity matching would also be an interesting comparisons as an alternative approach which has slightly different assumptions
- Consideration of more scenarios
  - Any suggestions are welcome – please email or speak to me afterwards
- Publication – a chance to fully explain the approach, and limitations as I see them
  - The aim is to have a paper submitted by the end of the year

References

- Phillippo, D., Ades, A.E., Dias, S., Palmer, S., Abrams, K.R., Welton, N.J., 2016. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE.
- Signorovitch, J.E., Wu, E.Q., Andrew, P.Y., Gerrits, C.M., Kantor, E., Bao, Y., Gupta, S.R., Mulani, P.M., 2010. Comparative Effectiveness Without Head-to-Head Trials. Pharmacoeconomics 28, 935–945.

Sign up as a Special Interest Group Member

- Visit ISPOR webpage www.ispor.org
- Click Member Groups
- Select ISPOR Special Interest Groups
- Must be an ISPOR member
- For more information, e-mail sigs@ispor.org

ISPOR Special Interest Group