What Causal Inference Teaches us about the Limitations of Indirect Treatment Comparisons for Health Technology Assessment

Workshop, ISPOR 2025 Montreal, 05/14/2025:

Uwe Siebert, MD, MPH, MSc, ScD,
UMIT TIROL – University for Health Sciences and Technology, and Harvard University

Arthur Chatton, PhD,
Département de Médecine Sociale et Préventive, Université Laval

Michael Webster-Clark, Pharm D, PhD,
Wake Forest University School of Medicine

Harlan Campbell, PhD,
Department of Statistics, University of British Columbia, and Precision AQ

Indirect treatment comparisons (ITCs) are essential in HTA when direct head-to-head trials are unavailable. However, ITCs are susceptible to biases threatening their validity. This workshop will introduce attendees to the key principles of causal inference and treatment effect heterogeneity and explain how these principles can improve our understanding of the limitations of ITCs within the framework of HTA. Participants will learn why ITCs are "essentially observational findings across trials" (Cochrane Handbook) and how to critically evaluate their validity.

Motivation for the Use of ITC

- Direct (head-to-head) randomized clinical trials not always possible:
  - Unfeasible/impractical
  - Life-threatening conditions with high unmet need
  - Unethical to enroll patients to a placebo control arm
  - Observational studies needed => causal thinking
- HTA requires indirect treatment comparisons
  - RCTs usually don't have all treatment arms desired for HTA
  - In the absence of direct RCT comparisons versus all candidate comparators, indirect treatment comparisons (ITCs) across studies are required.
  - ITCs are "essentially observational findings across trials, and may suffer the biases of observational studies, for example due to confounding" [Cochrane]

  => ITCs need causal thinking

Speakers

Moderator	Disclosures
Uwe Siebert, MPH, MSc, ScD, MD UMIT TIROL, Hall in Tirol, Austria & Harvard University, Boston, USA	No conflicts of interest relevant to the content of this workshop
Speakers	Disclosures
Arthur Chatton, PhD Faculté de Médecine, Université Laval, Québec, Québec, Canada	No conflicts of interest relevant to the content of this workshop
Michael Webster-Clark, PharmD, PhD Wake Forest University School of Medicine Winston-Salem, North Carolina, USA	No conflicts of interest relevant to the content of this workshop
Harlan Campbell, PhD Precision AQ & University of British Columbia, Vancouver, British Columbia, Canada	No conflicts of interest relevant to the content of this workshop

Overview

- **Uwe Siebert:** Introduction to key principles of causality, causal diagrams and methods.
- **Arthur Chatton:** Formal definition of causal effects using potential-outcomes framework; assumptions; types of causal estimators for SAT+ECA.
- **Michael Webster-Clark:** Role of effect measure, also determines the set of variables to be controlled for.
- **Harlan Campbell:** How principles translate to ITC; transportability; different estimands; adjustments.

Principles of Causality, Causal Diagrams and Causal Methods

Workshop, ISPOR 2025 Montreal, 05/14/2025: What Causal Inference Teaches Us About the Limitations of Indirect Treatment Comparisons for Health Technology Assessment

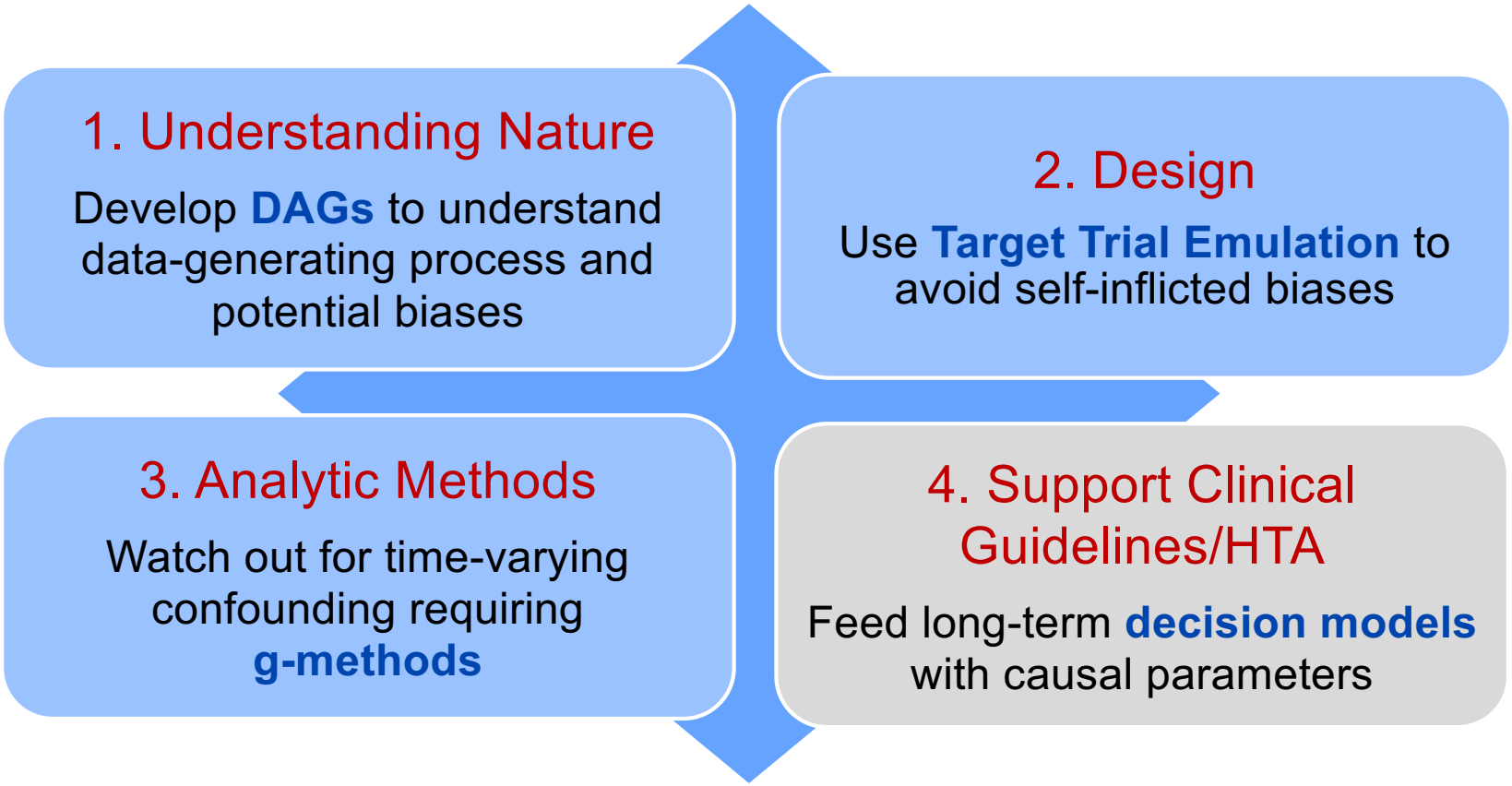
Uwe Siebert, MD, MPH, MSc, ScD

UMIT TIROL - University for Health Sciences and Technology, Hall in Tirol, Austria
& Harvard Chan School of Public Health, Boston, USA

Acknowledgment

This presentation was supported by the
UMIT TIROL Continuing Education Program
www.htads.org

4 Key Elements of a Causal Health Decision Framework



Causal Diagrams
(Directed Acyclic Graphs, DAGs)

Tx: Treatment
OC: Outcome
L: Covariate

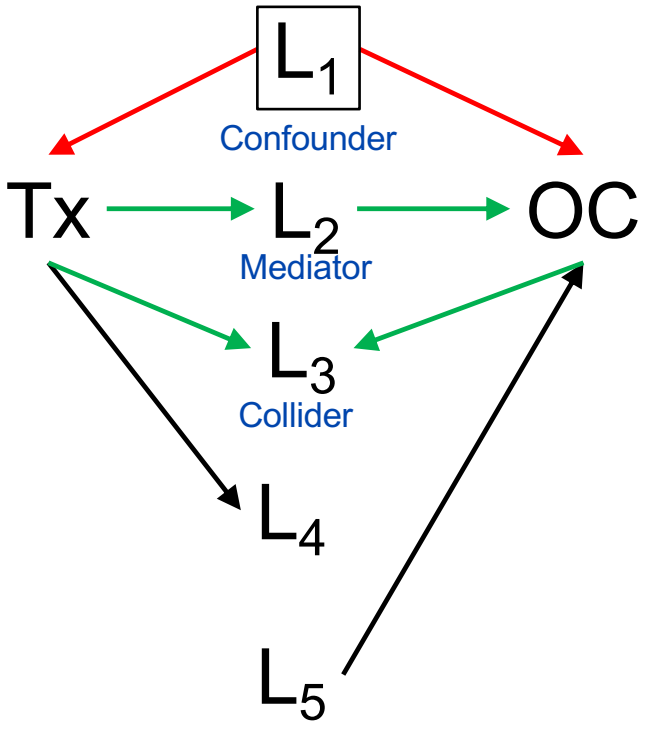
Graph is directed (arrows) and acyclic (no loops)

The total statistical association is represented by the sum of all open paths

There are frontdoor paths and backdoor paths

In the analysis, we must adjust (control) for open backdoor paths, to remove non-causal association (confounding)

Different types of variables …



Causal Diagrams
(Directed Acyclic Graphs, DAGs)

Tx: Treatment
OC: Outcome
L: Covariate

Which variables should we control for?

Modern definition of confounding:

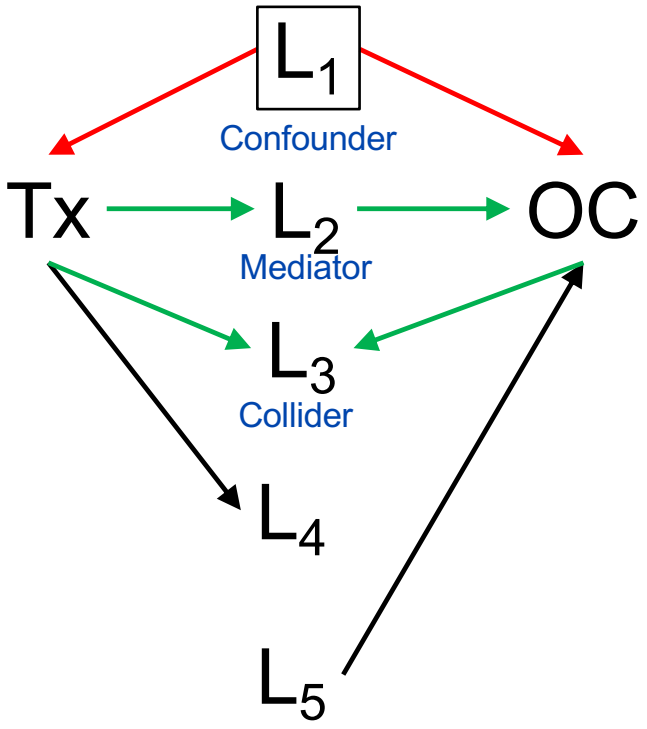
Open backdoor path

Adjustment:

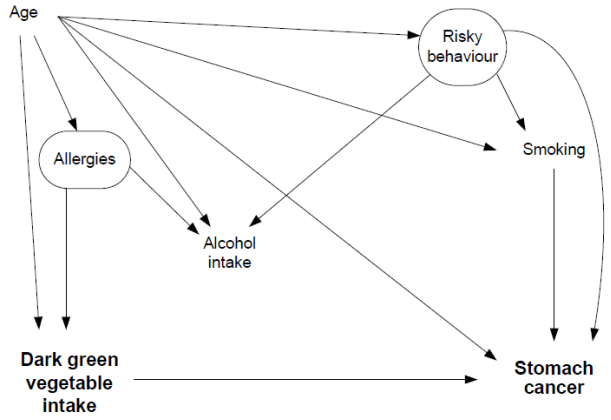
Block (control/adjust for) all open backdoor paths

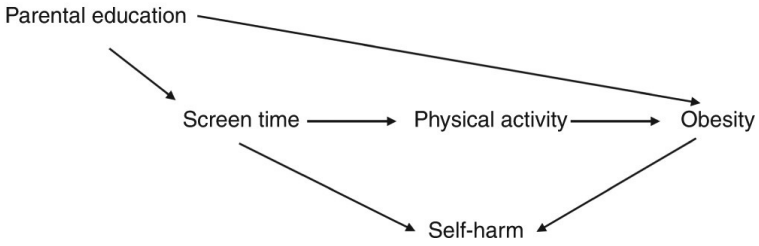
Rule:

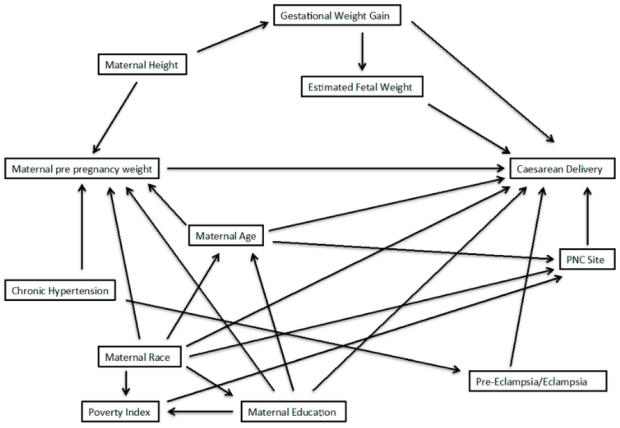
Never control for the future of the treatment

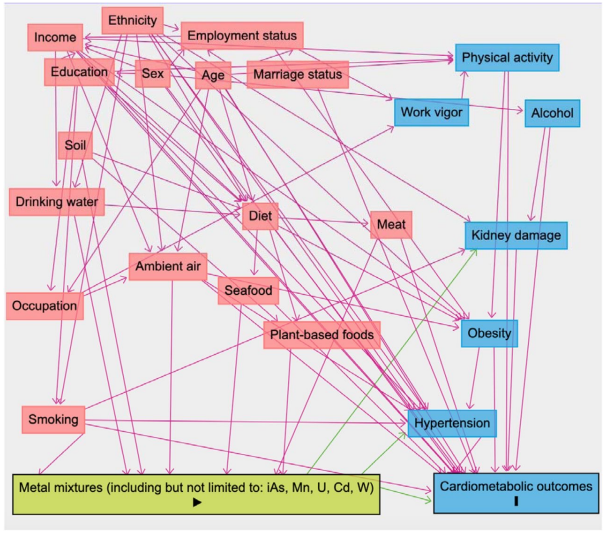


DAG Examples from the Literature









Source: Internet

Target Trial Emulation

Target Trial Approach

American Journal of Epidemiology
© The Author 2016. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 183, No. 8
DOI: 10.1093/aje/kwv254
Advance Access publication: March 18, 2016

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

Initially submitted December 9, 2014; accepted for publication September 8, 2015.

Design an observational study as if it was a randomized controlled experiment → develop protocol for a hypothetical RCT

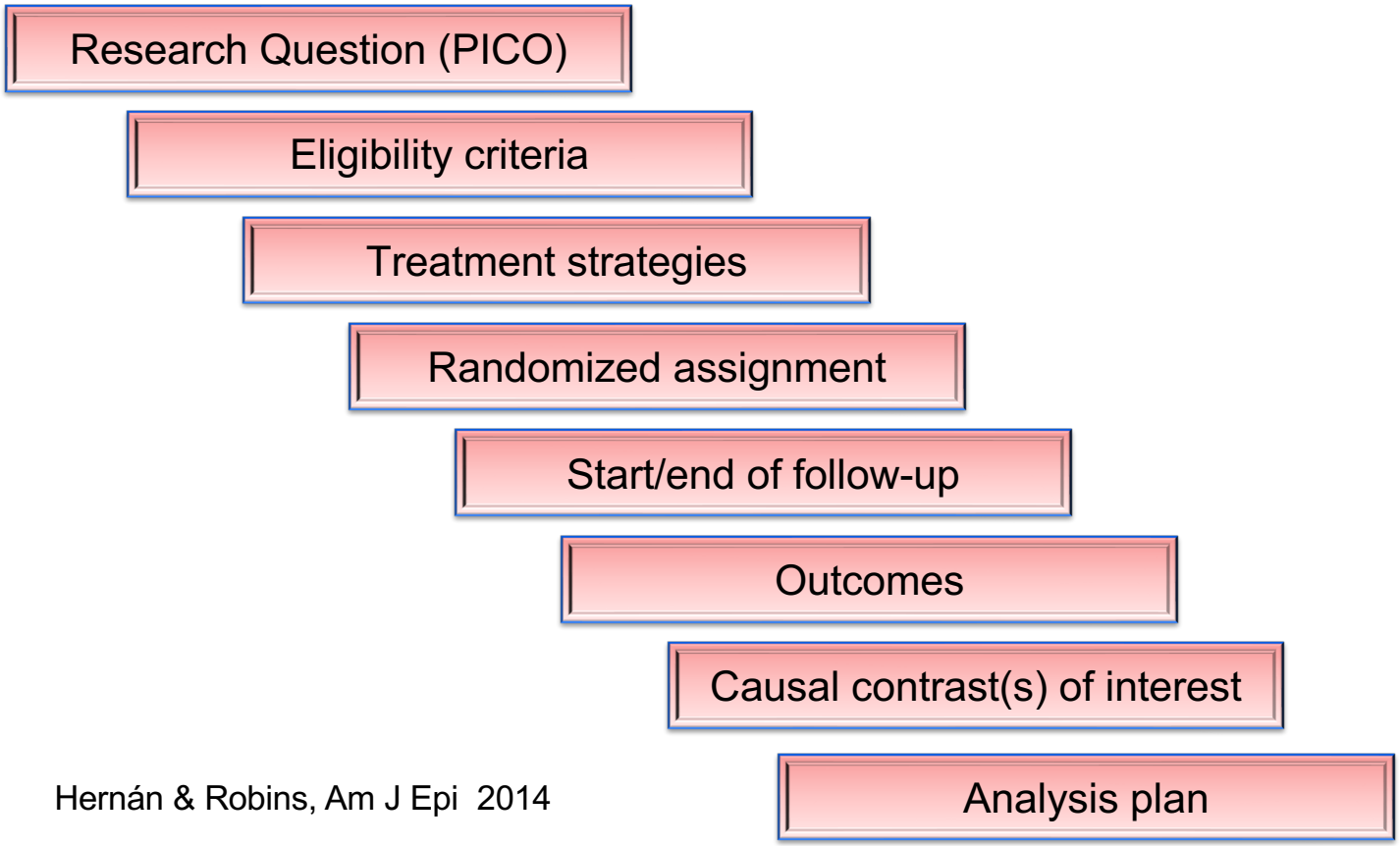
"Do not look into the future"

By defining all steps, the potential of self-inflicted biases (time-related biases, selection bias) is reduced

Time zero (time of including patients (and data), duration of follow up, etc.

Example with bias assessment see: Kuehne et al., JCE 2022

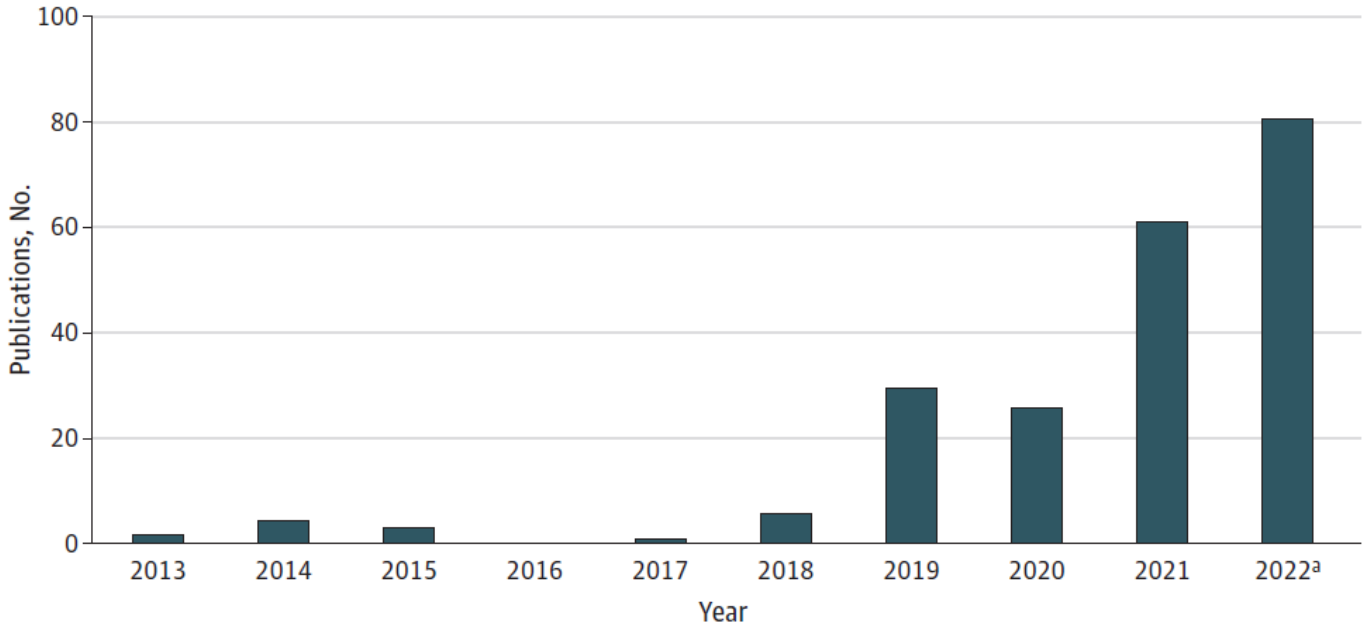
Target Trial Study Protocol



Hernán & Robins, Am J Epi 2014

Publications with Target Trial Emulation

Figure 2. Number of Explicit Emulations of a Target Trial Included in Review Published per Year

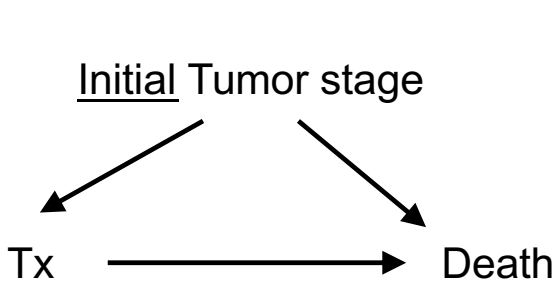


Hansford HJ et al., JAMA Network Open. 2023
https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2809945

Causal Inference Methods

Motivation: Confounding

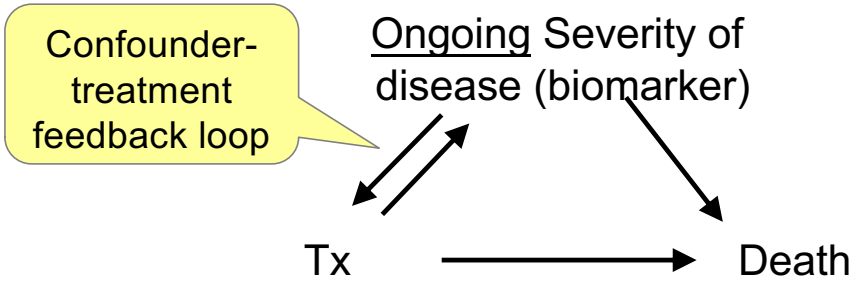
Tx = Treatment



Time-independent confounding
= *"baseline confounding"*

Initial Tumor stage is a common cause of prescribed Tx and Death

Traditional methods **work** (stratification, multivariable regression, propensity score etc.)



Time-dependent confounding
= *"post-baseline confounding"*

Severity is a common cause of Tx and Death and is also affected by Tx

Traditional methods **fail**,
→ apply **g-methods**

Quantitative Methods to Control for Confounding

Time-independent (Baseline) Confounding

- Traditional **methods**
  - Restriction
  - Stratification
  - Multivariate modeling
  - Matching
  - Propensity score
- [g-Methods]

Time-dependent (Post-baseline) Confounding

- g-Methods
  - g-formula
  - g-estimation
  - inverse probability weighting
- Further approaches:
  - Doubly robust methods (TMLE)

Causal Modeling

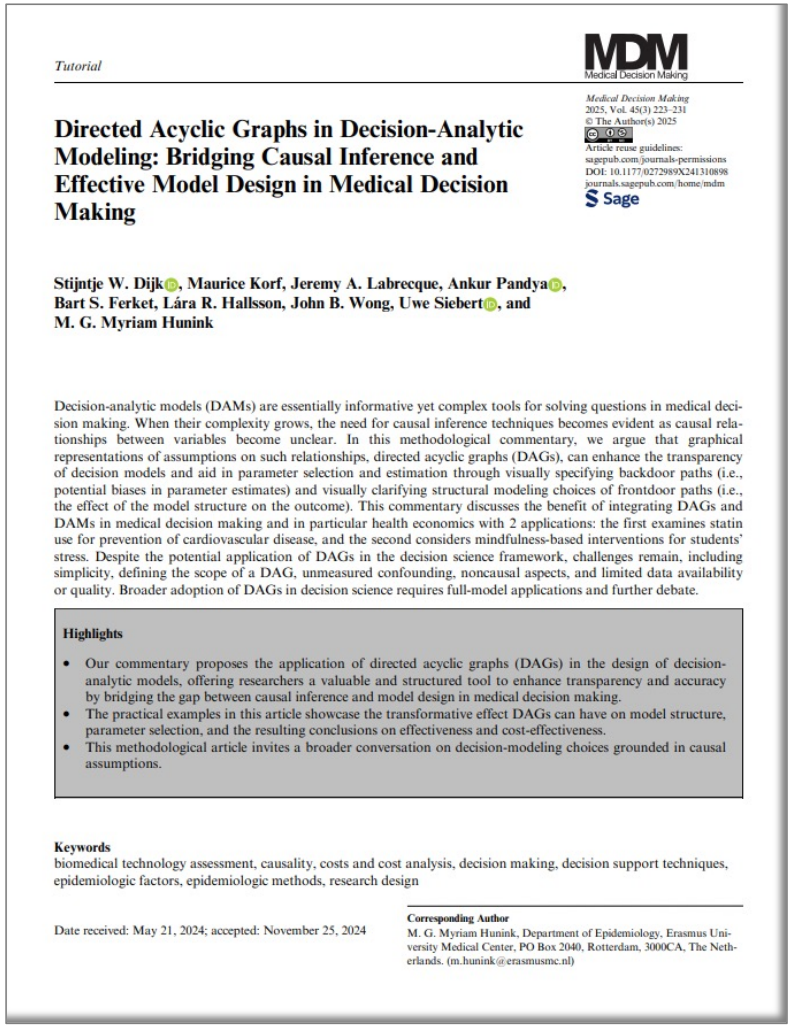
Causal Diagrams Informing Decision Models

Causal diagrams describe the causal relations between variables. We can use them to (1) build a causal natural history model and (2) to inform methods of empirical data analysis

Read more about how to match decision-analytic models with causal diagrams in this paper …

Dijk SW, Korf M, Labrecque JA, Pandya A, Ferket BS, Hallsson LR, Wong JB, Siebert U, Hunink MGM. Directed Acyclic Graphs in Decision-Analytic Modeling: Bridging Causal Inference and Effective Model Design in Medical Decision Making. Med Decis Making. 2025 Apr;45(3):223-231. doi: 10.1177/0272989X241310898.

New!



Tutorial

MDM Medical Decision Making

Medical Decision Making
2025, Vol. 45(3) 223–231
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X241310898
journals.sagepub.com/home/mdm
Sage

Directed Acyclic Graphs in Decision-Analytic Modeling: Bridging Causal Inference and Effective Model Design in Medical Decision Making

Stijntje W. Dijk, Maurice Korf, Jeremy A. Labrecque, Ankur Pandya, Bart S. Ferket, Lára R. Hallsson, John B. Wong, Uwe Siebert, and M. G. Myriam Hunink

Decision-analytic models (DAMs) are essentially informative yet complex tools for solving questions in medical decision making. When their complexity grows, the need for causal inference techniques becomes evident as causal relationships between variables become unclear. In this methodological commentary, we argue that graphical representations of assumptions on such relationships, directed acyclic graphs (DAGs), can enhance the transparency of decision models and aid in parameter selection and estimation through visually specifying backdoor paths (i.e., potential biases in parameter estimates) and visually clarifying structural modeling choices of frontdoor paths (i.e., the effect of the model structure on the outcome). This commentary discusses the benefit of integrating DAGs and DAMs in medical decision making and in particular health economics with 2 applications: the first examines statin use for prevention of cardiovascular disease, and the second considers mindfulness-based interventions for students' stress. Despite the potential application of DAGs in the decision science framework, challenges remain, including simplicity, defining the scope of a DAG, unmeasured confounding, noncausal aspects, and limited data availability or quality. Broader adoption of DAGs in decision science requires full-model applications and further debate.

Highlights

- Our commentary proposes the application of directed acyclic graphs (DAGs) in the design of decision-analytic models, offering researchers a valuable and structured tool to enhance transparency and accuracy by bridging the gap between causal inference and model design in medical decision making.
- The practical examples in this article showcase the transformative effect DAGs can have on model structure, parameter selection, and the resulting conclusions on effectiveness and cost-effectiveness.
- This methodological article invites a broader conversation on decision-modeling choices grounded in causal assumptions.

Keywords
biomedical technology assessment, causality, costs and cost analysis, decision making, decision support techniques, epidemiologic factors, epidemiologic methods, research design

Date received: May 21, 2024; accepted: November 25, 2024

Corresponding Author
M. G. Myriam Hunink, Department of Epidemiology, Erasmus University Medical Center, PO Box 2040, Rotterdam, 3000CA, The Netherlands. (m.hunink@erasmusmc.nl)

UMIT TIROL

Causal Inference in Indirect Treatment

- Make sure the research question is well defined
- Draw your DAG
- Perform a target trial emulation
- Use the correct estimators as parameters in your health-economic model

Questions? Contact:

@uweSiebert9

uwe-siebert9

uwe.siebert@umit-tirol.at

Web: www.htads.org

Causal effects and causal estimators

Arthur Chatton

Faculté de Médecine, Université Laval
CRM-StatLab-CANSSI postdoctoral fellow
arthur.chatton.1@ulaval.ca

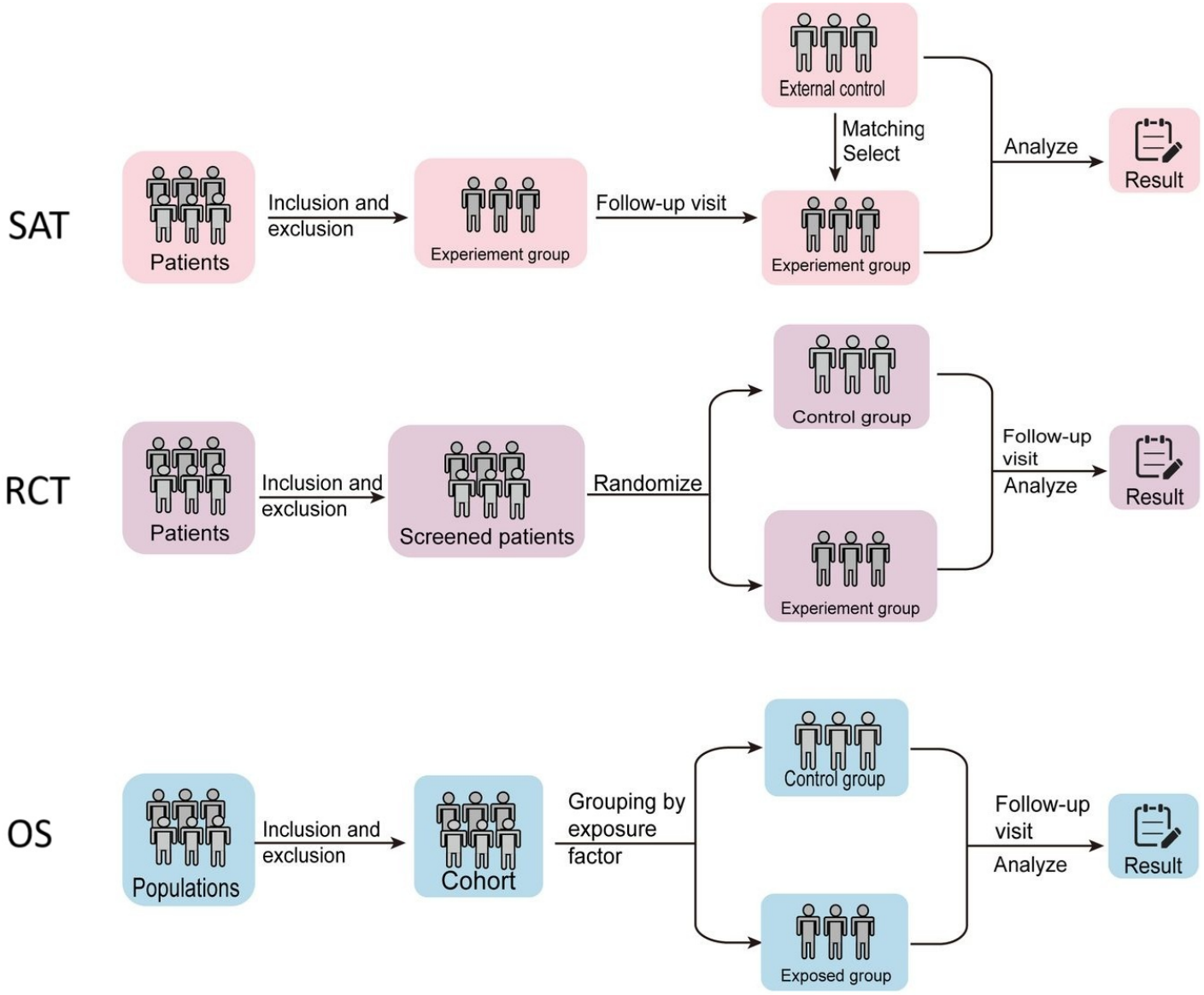
ISPOR – 14 May 2025

Goals

To gain a general understanding of causal inference

1. Counterfactual framework
2. What is a causal effect
3. Propensity score
4. G-computation
5. Doubly robust estimator

Single arm trials



Adapted from Wang et al. (2025)

Counterfactual framework

Causal inference: Quantifying the effect of the treatment A on the outcome Y

But what does the "effect" mean?

$\Rightarrow$ **Estimand**

Counterfactual framework

Each individual had two **potential outcomes** (assuming a binary A):

1. $Y^{a=1}$: Outcome observed in a hypothetical world where all are treated (A=1)
2. $Y^{a=0}$: Outcome observed in a hypothetical world where all are **un**treated (A=0)

Only one is actually observed

A	Y	Y^1	Y^0
0	1	?	1
1	1	1	?

Counterfactual framework

Each individual had two **potential outcomes** (assuming a binary A):

1. $Y^{a=1}$: Outcome observed in a hypothetical world where all are treated (A=1)
2. $Y^{a=0}$: Outcome observed in a hypothetical world where all are **un**treated (A=0)

Only one is actually observed

A	Y	Y^1	Y^0
0	1	?	1
1	1	1	?

Estimand = (Causal effect = contrast between Y^1 and Y^0) in a defined population

- Examples: $\underbrace{\mathrm{E}(Y^{a=1}) - \mathrm{E}(Y^{a=0})}_{\mathrm{ATE}}$ or $\underbrace{\mathrm{E}(Y^{a=1}|A=1)/\mathrm{E}(Y^{a=0}|A=1)}_{\mathrm{ATT}}$

$\Rightarrow$ Causal effect is based on **unmeasurable** variables

Counterfactual framework

Causal effect = concept, e.g., $\mathrm{E}(Y^1) - \mathrm{E}(Y^0)$

$\neq$

Association = Statistical measure, e.g., $\mathrm{E}(Y|A = 1] - \mathrm{E}(Y|A = 0)$

Assumptions are needed to map these two quantities $\Rightarrow$ **Identifiability**

Counterfactual framework

Causal effect = concept, e.g., $\mathrm{E}(Y^1) - \mathrm{E}(Y^0)$

$\neq$

Association = Statistical measure, e.g., $\mathrm{E}(Y|A = 1] - \mathrm{E}(Y|A = 0)$

Assumptions are needed to map these two quantities $\Rightarrow$ **Identifiability**

Identifiability

1. Consistency: $A_i = a \Rightarrow Y_i = Y_i^a$
2. Exchangeability: $Y^a \perp\!\!\!\perp A|X$
3. Positivity: $P(A = a|X = x) > \beta \geq 0, \forall x$ such as $P(X = x) > 0$

- ITT: main estimand in RCT
- SATs usually target the Per Protocol effect
  - $\Rightarrow$ contrast between $\mathrm{E}(Y^{a=1,\bar{c}=1})$ and $\mathrm{E}(Y^{a=0,\bar{c}=1})$
    - Identifiability assumptions must also apply for the censoring history $\bar{C}$
- The external control group should be chosen to minimize potential consistency violations
- Residual confounding is likely $\Rightarrow$ Adjusting on some X

Counterfactual framework Adjusting on X

- FDA (2019) and EMA (2015) recommend adjusting on a few X prognostics of Y to increase the statistical power
- This targets the Conditional Average Treatment Effect: $\mathrm{E}(Y^1|X=x) - \mathrm{E}(Y^0|X=x)$
  - Not really useful for non-collapsible measures, such as the OR/HR
- Causal estimators adjust on X and recover the ATE (RCT goal)
- Can be used in SAT to control residual confounding too

Causal estimators

Three “families” of causal estimators:

1. Propensity score (PS): Model A conditional on X to balance the arms
2. G-computation (or g-formula): Model Y conditional on A and X to predict Y^a
3. Doubly robust estimators (DRE): Combine PS and g-computation to reduce modelling assumptions

I will not speak about quasi-experimental approaches, such as IV, today.

Causal estimators Propensity score

Goal of PS: balance the treatment arms to mimic an RCT

Can be used in various ways:

- Stratification on the PS
- Added as the covariate in a model
- Matching on the PS
- Weighting on the PS

General recipes:

1. Model $e(X) = P(A = 1|X)$ using for instance a logistic regression

Now for Matching:

2. Match treated and untreated individuals according to their $e(X_i)$
3. Model $\mathrm{E}(Y|A)$ on the resulting matched sample

General recipes:

1. Model $e(X) = P(A = 1|X)$ using for instance a logistic regression

Now for Matching:

2. Match treated and untreated individuals according to their $e(X_i)$
3. Model $\mathrm{E}(Y|A)$ on the resulting matched sample

While for Weighting (i.e., inverse probability weighting):

2. Compute the individuals' weight ω as a function of $e(X_i)$
3. Model $\mathrm{E}(Y|A)$ on the whole sample, but weighted by ω

The coefficient of A will be the causal effect (the type of regression should match the wanted contrast).
Use a sandwich estimator or bootstrap for the variance.

Causal estimators G-computation

Goal of g-computation: Simulate the counterfactual worlds

General recipe:

1. Model $Q(A, X) = \mathrm{E}(Y|A, C)$ using all individuals of the sample
2. Predict $Y^a = Q(a, X_i), \forall a$ for each individuals
3. Average Y^a over all individuals and contrast them to compute the estimand of interest

Causal estimators Pro and cons

- Positivity is directly checkable with PS, but g-computation may be more robust to random violations
- PS and g-computation require a correct specification of $e(X)$ or $Q(A, X)$, resp.
- DREs relax this by requiring the correct specification of one but not necessarily both
- They also allow the use of machine learning for fitting these models, further limiting the risk of misspecification bias
- Several DREs exist; I will focus on the easiest one (IMHO)

Causal estimators Doubly robust standardisation

Goal of doubly robust standardization: use PS weighting to mimic RCT, then use g-computation to remove residual confounding

General recipe:

- Model $e(X) = P(A = 1|X)$ using for instance a logistic regression
- Compute the individuals' weight ω as a function of $e(X_i)$
- Model $Q(A, X) = \mathrm{E}(Y|A, X)$ using all individuals of the sample, **but the model is weighted by** ω
- Predict $Y^a = Q(a, X_i), \forall a$ for each individuals
- Average Y^a over all individuals and contrast them to compute the estimand of interest

Takeaways

- Residual confounding in SAT can be controlled if measured
- Causal estimators are easily computable using standard statistical software
- Quality of data is still crucial

Takeaways

- Residual confounding in SAT can be controlled if measured
- Causal estimators are easily computable using standard statistical software
- Quality of data is still crucial

To go further:

$\Rightarrow$ Chatton and Rohrer (2024)

Other readings:

- Review on matching schemes: Stuart (2010)
- Tutorials on DREs:
  - Binary outcome: Luque-Fernandez et al. (2018)
  - Time-to-event outcome: Talbot et al. (2025)

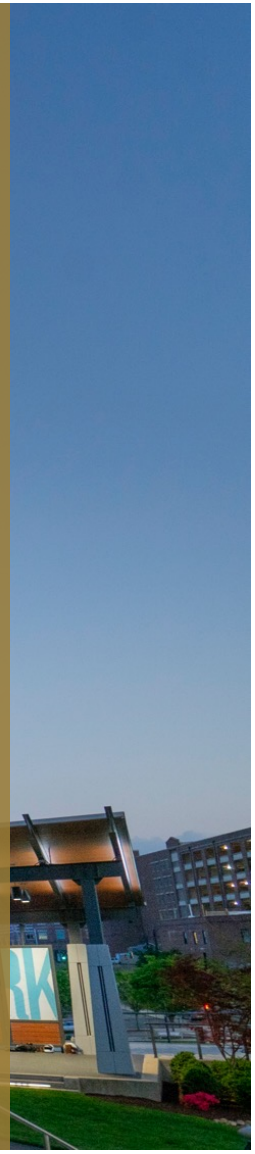
Bibliography

- Chatton A, Rohrer JM. The Causal Cookbook: Recipes for Propensity Scores, G-Computation, and Doubly Robust Standardization. Advances in Methods and Practices in Psychological Science. 2024;7(1):25152459241236149.
- EMA. Guideline on adjustment for baseline covariates in clinical trials. 2015. Report No.: EMA/CHMP/295050/2013.
- FDA. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products: Guidance for Industry. 2023. Report No.: FDA-2019-D-0934.
- Luque-Fernandez MA, Schomaker M, Rachet B, Schnitzer ME. Targeted maximum likelihood estimation for a binary treatment: A tutorial. Stat Med. 2018;37(16):2530–46.
- Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1–21.
- Talbot D, Diop A, Mésidor M, Chiu Y, Sirois C, Spieker AJ, et al. Guidelines and Best Practices for the Use of Targeted Maximum Likelihood and Machine Learning When Estimating Causal Effects of Exposures on Time-To-Event Outcomes. Stat Med. 2025;44(6):e70034.
- Wang M, Ma H, Shi Y, Ni H, Qin C, Ji C. Single-arm clinical trials: design, ethics, principles. BMJ Supportive & Palliative Care. 2025;15(1):46–54.


How Effect Measure Elements Impact External Validity
ISPOR 2025
MAY 14TH, 2025
MICHAEL WEBSTER-CLARK, PHARM D, PHD
Wake Forest University School of Medicine
The academic core of
Atrium Health

Table of contents

What’s in an effect measure?



Two effect measure descriptions

- What is the effect of dabigatran on gastrointestinal bleeding?
- What is the 2-year risk difference for gastrointestinal bleeding for Medicare beneficiaries recently diagnosed with atrial fibrillation comparing initiating and staying on treatment for atrial fibrillation with dabigatran (vs warfarin)?
- Which of these…
  - Is a well-defined causal effect?
  - Could fit in the title of a journal article?

What makes for a well-defined causal effect?

- Somewhat obviously…
  - Treatment of interest: dabigatran
  - Referent or comparator: warfarin
  - Outcome: mortality
  - Intention-to-treat vs per-protocol: initiating and staying on treatment
- More subtly…
  - The index date: recent diagnosis with atrial fibrillation
  - The target population(s): Medicare beneficiaries with atrial fibrillation
  - Scale of the effect: the risk difference
  - Length of follow-up: 2 years
  - Handling of competing events: allowed to occur (should be the default)

Deciding which effect to estimate

- Should always consider…
  - Stakeholder opinions
  - Potential for intractable confounding or selection bias
  - Practical issues with data availability
  - How they impact external validity (…hold that thought)
- Like almost everything in public health, this is a **balancing act**

Defining external validity

Crude external validity

- The effect in the study population equals the effect in the target population
- For the RD…
  - $E(Y^{a=1}|Study) - E(Y^{a=0}|Study) = E(Y^{a=1}|Target) - E(Y^{a=0}|Target)$
- Achieved via…
  - Random sampling
  - Homogeneity (if collapsible measure)

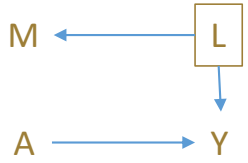
What is effect measure modification?

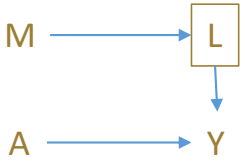
- Effect measure modification by a variable M occurs when the **effect measure of interest** varies across levels of M
- Mathematically,
  - $E(Y^{a=1}|EMM=1) - E(Y^{a=0}|EMM=1) \neq E(Y^{a=1}|EMM=0) - E(Y^{a=0}|EMM=0)$
- Several different types of effect measure modification, all threatening external validity

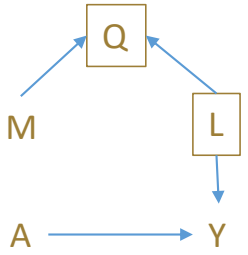
- If M represents "population", we lack crude external validity **on either the RD or RR scale** in all of these scenarios

Conditional external validity

- ***Conditional on an adjustment set L,*** the effect measure from the study population equals the effect measure in the target population
- For the RD:
  - $E(Y^{a=1}|Study,L=t) - E(Y^{a=0}|Study,L=t) = E(Y^{a=1}|Target) - E(Y^{a=0}|Target)$
- Where $L=t$ means that we have adjusted for L via:
  - Weighting
  - The g formula and outcome modeling
  - Doubly robust approaches



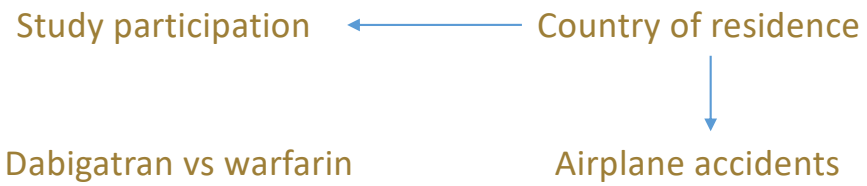


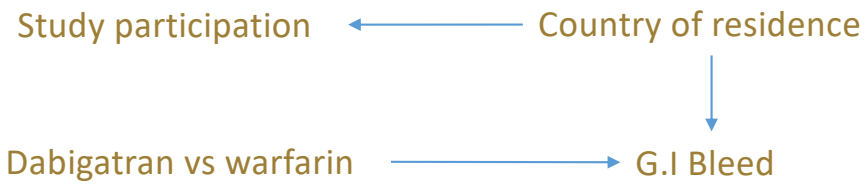


Different elements of effect measures and external validity

The outcome and follow-up duration

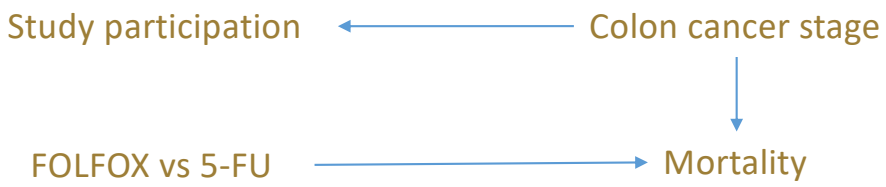
- Changing the outcome can directly impact external validity
  - (for pretty obvious reasons)
- Outcomes where the treatment truly has no effect will always have crude external validity
- This can apply to short-term outcomes for treatments with long latency periods, as well





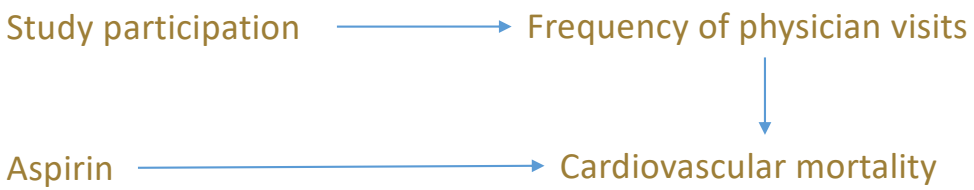
Index date

- Some treatments vary in safety/effectiveness over time
  - Thrombolytics for stroke prevention
  - Different types of rehabilitation
  - Treatments for Parkinson's
  - Everything in oncology
- This often manifests as effect measure modification, especially in the case of **critical windows**, and may need to be accounted for



The treatment and comparator

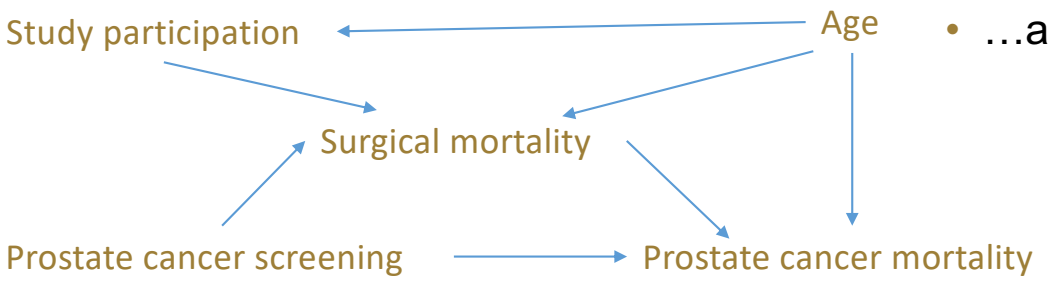
- Variations between study and target population interventions can threaten external validity due to issues with **consistency**
- Consider…
  - Financial incentives (or support) in the study that aren't present in the target
  - Differences in surveillance or lifestyle interventions
  - Actual differences between formulations (e.g., extended release)
- Can sometimes be fixed if an intermediate variable has been measured



Handling of competing events

Option 1: Censoring

- Censoring individuals who experience competing events estimates an effect where you can prevent all those events
- When populations differ in their risk of the competing event, this makes achieving external validity **easier**

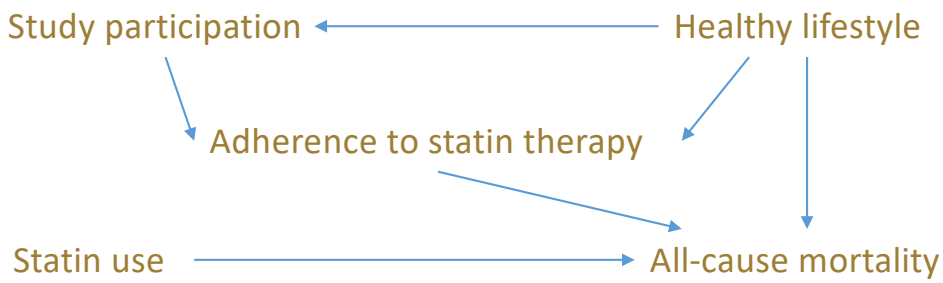


Option 2: Allowing to occur

- Allowing competing events to occur estimates a much **simpler to describe** causal effect
- You may also need to address differences in the risk of the competing event between populations, however
- …and it complicates analyses

Intention-to-treat vs per-protocol follow-up

- Intention-to-treat effects face external validity issues if adherence and **predictors of adherence differ** between populations
  - This can create additional effect measure modification
- Per-protocol effects **condition away** this source of potential effect measure modification by default
  - But will need to address potential informative censoring resulting from deviations from the protocol



Scale of the effect

- The scale of the effect has implications for adjustment sets
- The risk difference (RD) is a **weighted average** of each RD in the population, with weights based on the prevalence
  - Requires variables that directly interact with treatment on the difference scale, rather than overall/marginal EMMs
- The risk ratio (RR) is a **weighted average** of the RRs, with weights based on the prevalence **and** the risk of the outcome
  - Requires variables that interact on the ratio scale **and** variables that don't interact but are associated with the outcome
- The odds ratio (OR) isn't a weighted average of anything
  - Requires all variables associated with the outcome


Conclusion

External validity is complex

- Every decision you make about the effect measure you are estimating has implications for its external validity
- Conditions to achieve external validity are as complicated, if not more complicated, than conditions to achieve internal validity
- Sometimes, the safe choice for internal validity is not the safe choice for external validity
  - Placebo groups?
  - Intention-to-treat analyses?
- Thinking through all of these elements is a major part of generalizing or transporting effect estimates

A closing question

- Which do you prioritize when designing your studies and choosing an effect measure?

A. Internal validity

B. External validity

C. Overall validity

Questions

My emails: miwebste@wakehealth.edu, michael.webster-clark@mcgill.ca, mawc@live.unc.edu

Adjustment for external vs internal validity

Internal validity

- Weight the **exposed** and **unexposed** to a common population (ATE, ATT, ATU) based on predicted **treatment** probability

- Use outcome models built in the **exposed** and **unexposed** to predict outcomes in a common population

- Combine **treatment** and **outcome** models to obtain doubly-robust estimates

External validity

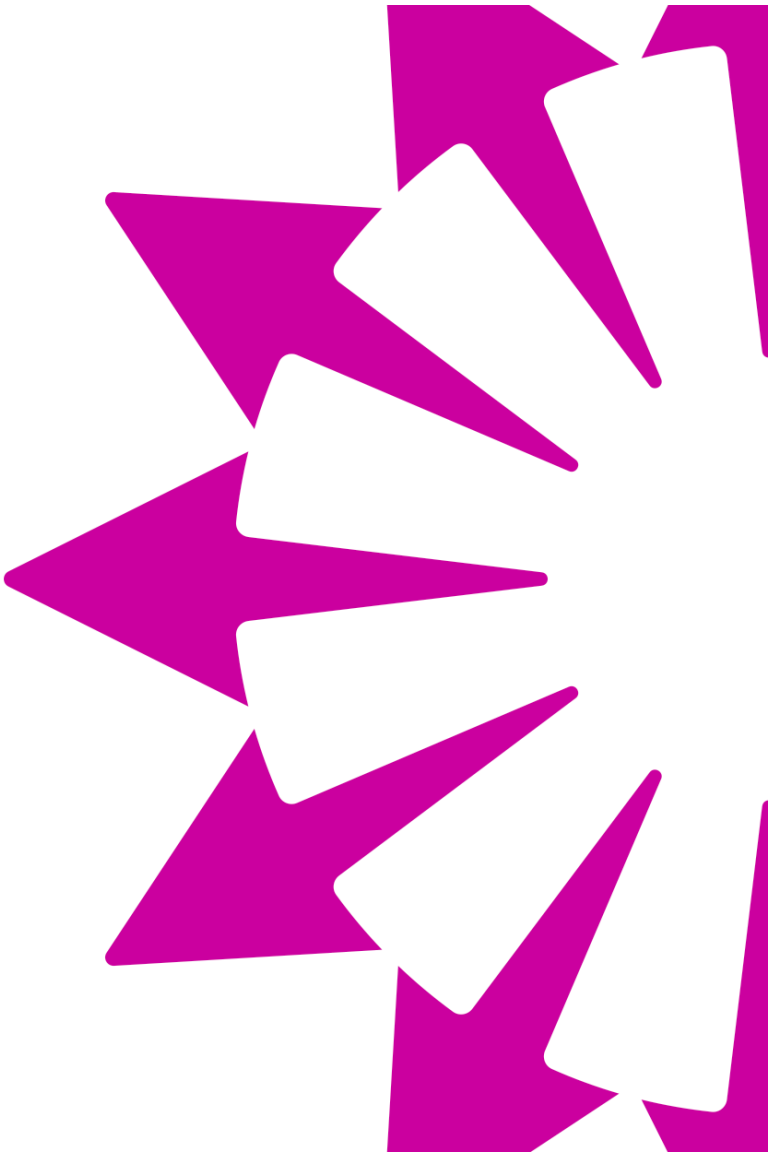
- Weight the **study population** to resemble the **target population of interest** based on predicted **“sampling”** probability
- Use outcome models built in the **study population** to predict outcomes in the **target population of interest**
- Combine **sampling** and **outcome** models to obtain doubly-robust estimates

What Causal Inference Teaches us about the Limitations of Indirect Treatment Comparisons for Health Technology Assessment

ISPOR 2025, Montreal

Harlan Campbell, PhD

Department of Statistics, University of British Columbia, and Precision AQ



A question

When conducting ITCs for HTA, why is an understanding of causal inference *most* important?

A. …to select the best estimator for the analysis.

B. …to ensure we are adjusting for all the necessary variables for unbiased estimation.

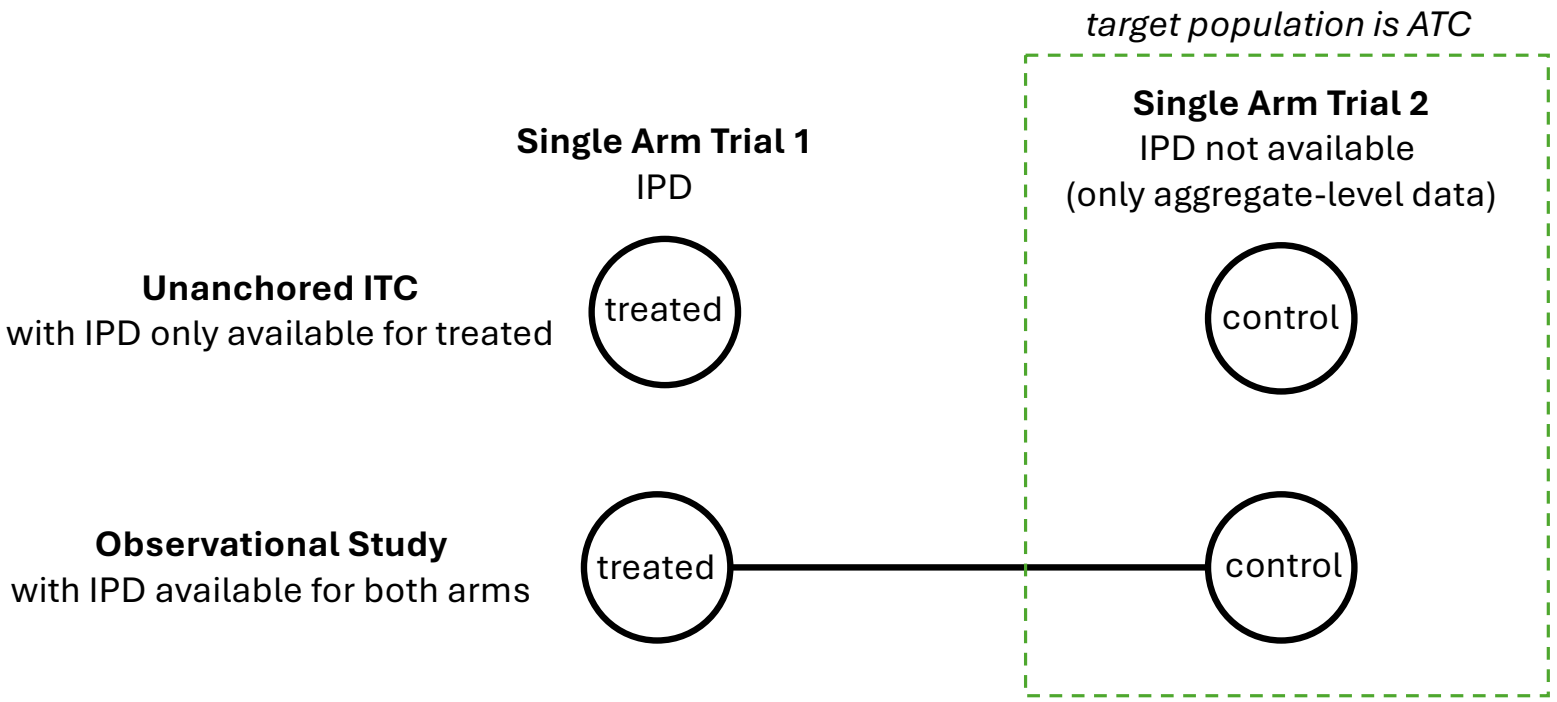
C. …to ensure we are targeting the correct estimand of interest.

ITCs for HTA require causal thinking: 3 examples

- **Example 1: An unanchored ITC**
  - *How do we adjust for confounders when IPD is unavailable?*
    - *MAIC or STC?*
    - *or augmented MAIC for doubly-robust estimation?*
- **Example 2: An anchored ITC**
  - *Different effect measure scales require adjusting for different variables.*
    - *Non-collapsible effect measures depend on the distribution of purely prognostic factors in the study population.*
    - *But the bias from ignoring purely prognostic factors will typically be small.*
- **Example 3: A three study NMA**
  - *What is the causal estimand of interest?*
    - *Marginal or population-average conditional?*

EXAMPLE 1

Unanchored Indirect Treatment Comparison without full IPD



EXAMPLE 1

Doubly robust augmented weighting estimators for the analysis of externally controlled single-arm trials and unanchored indirect treatment comparisons

Harlan Campbell[1,2] | Antonio Remiro-Azócar[3]

[1]Evidence Synthesis and Decision Modeling, Precision AQ, British Columbia, Canada
[2]Department of Statistics, University of British Columbia, British Columbia, Canada
[3]Methods and Outreach, Novo Nordisk Pharma, Madrid, Spain

Externally controlled single-arm trials are critical to assess treatment efficacy across therapeutic indications for which randomized controlled trials are not feasible. A closely-related research design, the unanchored indirect treatment comparison, is often required for disconnected treatment networks in health technology assessment. We present a unified causal inference framework for both research designs. We develop a novel estimator that augments a popular weighting approach based on entropy balancing – matching-adjusted indirect comparison (MAIC) – by fitting a model for the conditional outcome expectation. The predictions of the outcome model are combined with the entropy balancing MAIC weights. While the standard MAIC estimator is singly robust where the outcome model is non-linear, our augmented MAIC approach is doubly robust, providing increased robustness against model misspecification. This is demonstrated in a simulation study with binary outcomes and a logistic outcome model, where the augmented estimator demonstrates its doubly robust property, while exhibiting higher precision than all non-augmented weighting estimators and near-identical precision to G-computation. We describe the extension of our estimator to the setting with unavailable individual participant data for the external control, illustrating it through an applied example. Our findings reinforce the understanding that entropy balancing-based approaches should be augmented to improve protection against bias and guarantee double robustness.

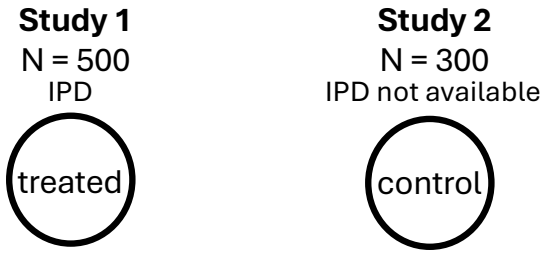
- For the ATC, “balancing” approaches to weighting (e.g., MAIC or entropy balancing), can enhance performance relative to standard “modeling” approaches (i.e., propensity score weighting).
- The augmented MAIC is doubly robust and has higher precision than non-augmented weighting estimators when the outcome model is correctly specified.
- The augmented MAIC achieves near-identical precision to G-computation/STC, which can have substantial bias where the outcome model is misspecified.

Campbell, H, Remiro-Azócar, A., (2025). Doubly robust augmented weighting estimators for the analysis of externally controlled single-arm trials and unanchored indirect treatment comparisons. *arXiv preprint*, arXiv:2505.00113

EXAMPLE 1

Different causal estimators require different assumptions

Unanchored Indirect Treatment Comparison



The target estimand is the effect of A vs B in the Study 2 population (i.e., the ATC) on the marginal log-odds ratio scale.

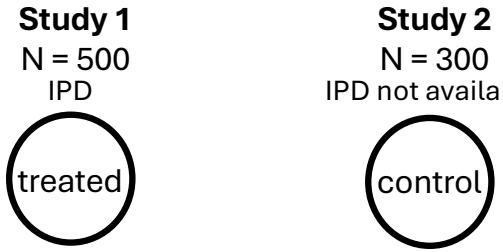
Four baseline covariates have been identified as potential confounders:

- age
- sex
- ECOG
- smoking status

Different causal estimators require different assumptions

EXAMPLE 1

Unanchored Indirect Treatment Comparison

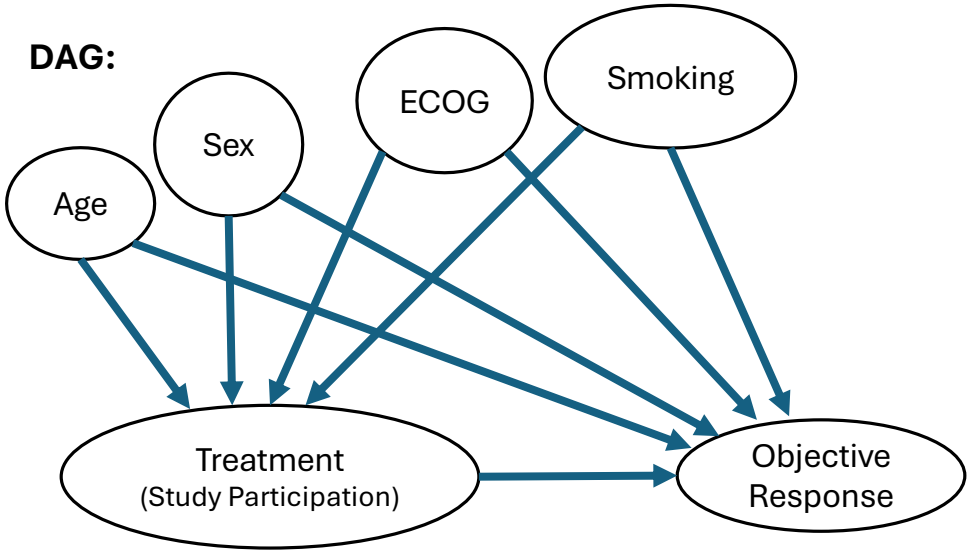


The target estimand is the effect of A vs B in the Study 2 population (i.e., the ATC) on the marginal log-odds ratio scale.

Four baseline covariates have been identified as potential confounders:

- age
- sex
- ECOG
- smoking status

DAG:



Covariate	Intervention SAT ($n_1 = 500$)	External control ($n_0 = 300$)
Age in years (mean; standard deviation)	59.85; 9.01	50.06; 3.24
Sex (proportion male)	0.38	0.49
ECOG (proportion ECOG performance status of 1)	0.41	0.35
Smoking status (proportion of smokers)	0.32	0.19

EXAMPLE 1

Three different estimators

- **Matching-adjusted indirect comparison (MAIC)**
  - *Weighting method similar to propensity score estimators*
  - ***Assumption:*** *Requires correct specification of the implied trial assignment model*

- **Simulated treatment comparison (STC)**
  - *Model-based standardization method similar to G-computation*
  - ***Assumption:*** *Requires correct specification of the outcome model*

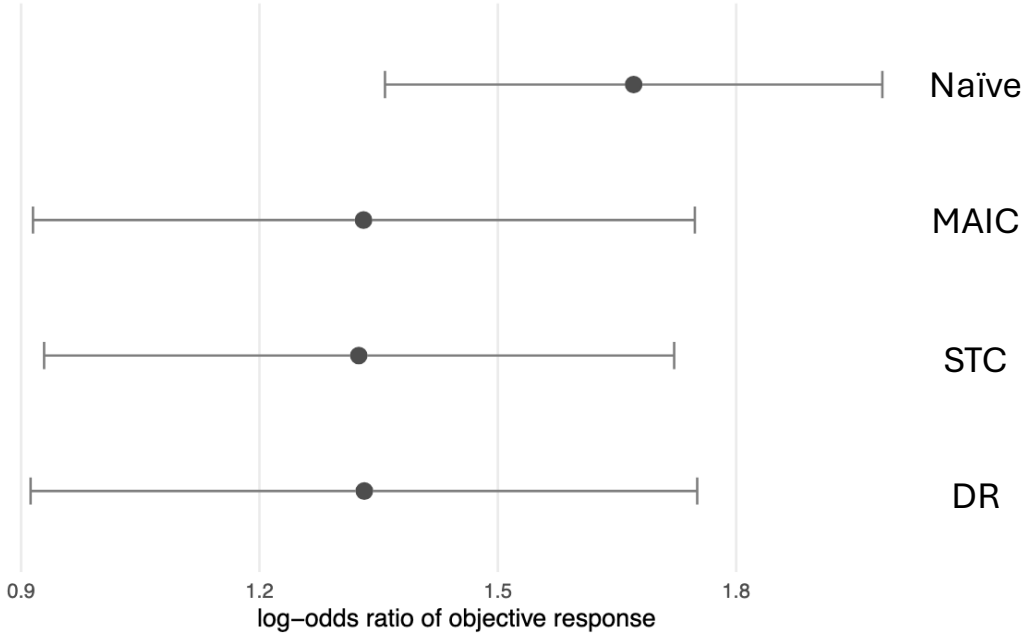
- **Doubly robust augmented MAIC estimator (DR)**
  - ***Assumption:*** *Requires either the propensity score model or the outcome model be correct, but not necessarily both*

EXAMPLE 1

Results

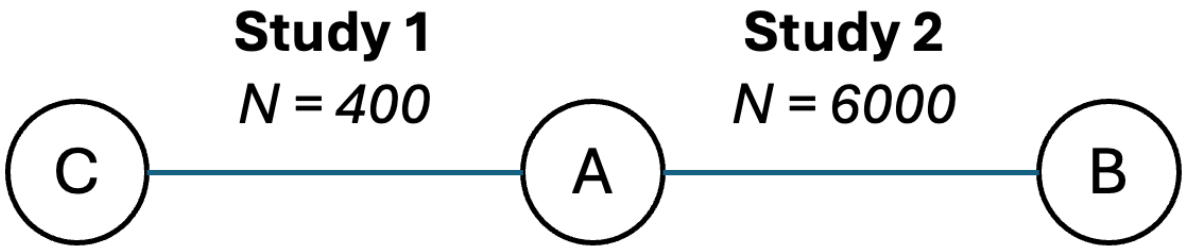
MAIC vs STC vs DR

- The intervention improves objective response versus the control.
- DR point estimate is not meaningfully different than the MAIC or STC point estimates.
- DR approach has slightly increased uncertainty than STC.
- This loss of precision seems a relatively minor price to pay for robustness to model misspecification.



EXAMPLE 2

Anchored Indirect Treatment Comparison



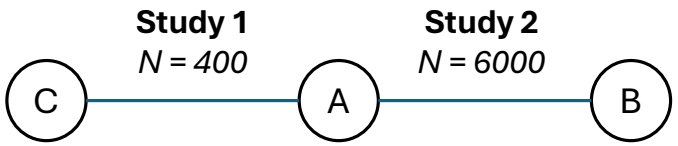
Two randomised trials with three treatments A, B, C, in populations where the prevalence of smoking varies substantially.

Riley, R. D., Dias, S., Donegan, S., Tierney, J. F., Stewart, L. A., Efthimiou, O., & Phillippo, D. M. (2023). Using individual participant data to improve network meta-analysis projects. *BMJ evidence-based medicine*, *28*(3), 197-203.

EXAMPLE 2

Transportability for Indirect Treatment Comparisons

Anchored Indirect Treatment Comparison

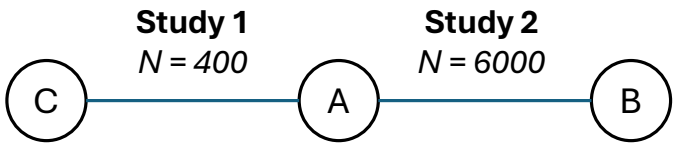


- Outcome of interest: the number of individuals who respond to treatment.
- IPD is available for Study 1 but not for Study 2.
- Smoking is prognostic but not effect-modifying (i.e., treatment is same for smokers and non-smokers).

EXAMPLE 2

Transportability for Indirect Treatment Comparisons

Anchored Indirect Treatment Comparison



- Outcome of interest: the number of individuals who respond to treatment.
- IPD is available for Study 1 but not for Study 2.
- Smoking is prognostic but not effect-modifying (i.e., treatment is same for smokers and non-smokers).

Objective: Estimate the relative treatment effect of *C vs B* in the Study 2 population

This requires ***transporting*** the effect of treatment C in the Study 1 population to the Study 2 population.

We need to think about external validity!

EXAMPLE 2

Study 1

Ratio of non-smokers to smokers is 1:1

Binary outcome (Response)	**Non-smokers** 		**Smokers** 		**Overall**	
	Response Yes	Response No	Response Yes	Response No	Response Yes	Response No
Treatment C	90	10	50	50	140	60
Treatment A	50	50	10	90	60	140
Risk difference: C/A	0.9 – 0.5 = 0.4		0.5 – 0.1 = 0.4		0.7 – 0.3 = 0.4	
Marginal odds ratio: C/A	$\frac{90/10}{50/50} = 9$		$\frac{50/50}{10/90} = 9$		$\frac{140/60}{60/140} = 5.4$	

EXAMPLE 2

Study 1

Ratio of non-smokers to smokers is 1:1

Binary outcome (Response)	**Non-smokers** 		**Smokers** 		**Overall**	
	Response Yes	Response No	Response Yes	Response No	Response Yes	Response No
Treatment C	90	10	50	50	140	60
Treatment A	50	50	10	90	60	140
Risk difference: C/A	0.9 – 0.5 = 0.4		0.5 – 0.1 = 0.4		0.7 – 0.3 = 0.4	
Marginal odds ratio: C/A	$\frac{90/10}{50/50} = 9$		$\frac{50/50}{10/90} = 9$		$\frac{140/60}{60/140} = 5.4$	

$9 = 9 \neq 5.4 \rightarrow$ Non-collapsibility!

EXAMPLE 2

Study 2

Ratio of non-smokers to smokers is 5:1

Binary outcome (Response)	**Non-smokers** 		**Smokers** 		**Overall**	
	Response Yes	Response No	Response Yes	Response No	Response Yes	Response No
Treatment B	2250	250	250	250	2500	500
Treatment A	1250	1250	50	450	1300	1700
Risk difference: B/A	0.9 – 0.5 = 0.4		0.5 – 0.1 = 0.4		0.83 – 0.43 = 0.4	
Marginal odds ratio: B/A	$\frac{2250/250}{1250/1250} = 9$		$\frac{250/250}{50/450} = 9$		$\frac{2500/500}{1300/1700} = 6.5$	

EXAMPLE 2

Study 2

Ratio of non-smokers to smokers is 5:1

Binary outcome (Response)	Non-smokers 		Smokers 		Overall	
	Response Yes	Response No	Response Yes	Response No	Response Yes	Response No
Treatment B	5000		1000		2500	500
Treatment A					1300	1700
Risk difference: B/A	0.83 – 0.43 = 0.4					
Marginal odds ratio: B/A	$\frac{2500/500}{1300/1700} = 6.5$					

EXAMPLE 2

Marginal Risk Difference of A vs. B in Study 2 population

Estimated Marginal Risk Difference:

$$\widehat{\mathrm{RD}}_{BC}^{S=2} = \widehat{\mathrm{RD}}_{AC}^{S=2} - \widehat{\mathrm{RD}}_{AB}^{S=2}$$

"Unadjusted" model

Ignoring data on smoking status

$$\widehat{\mathrm{RD}}_{BC}^{S=2} = 0.4 - 0.4 = 0.0$$

Conclusion: There is no difference between A and B in the Study 2 population

EXAMPLE 2

Marginal Odds Ratio of A vs. B in Study 2 population

Estimated Marginal Odds Ratio:

$$\log\left(\widehat{\mathrm{OR}}_{BC}^{S=2}\right) = \log\left(\widehat{\mathrm{OR}}_{AC}^{S=2}\right) - \log\left(\widehat{\mathrm{OR}}_{AB}^{S=2}\right)$$

"Unadjusted" estimator

Ignoring data on smoking status

$$\widehat{\mathrm{OR}}_{BC}^{S=2} = \exp(\log(5.4) - \log(6.5)) = 0.83$$

Conclusion: Treatment C is worse than treatment B in the Study 2 population

Population-adjusted estimators

MAIC

$$\widehat{\mathrm{OR}}_{BC}^{S=2} = \exp(\log(6.5) - \log(6.5)) = 1$$

Using data on smoking status

Assumption: *Requires correct specification of the implied propensity score model*

STC

$$\widehat{\mathrm{OR}}_{BC}^{S=2} = \exp(\log(6.5) - \log(6.5)) = 1$$

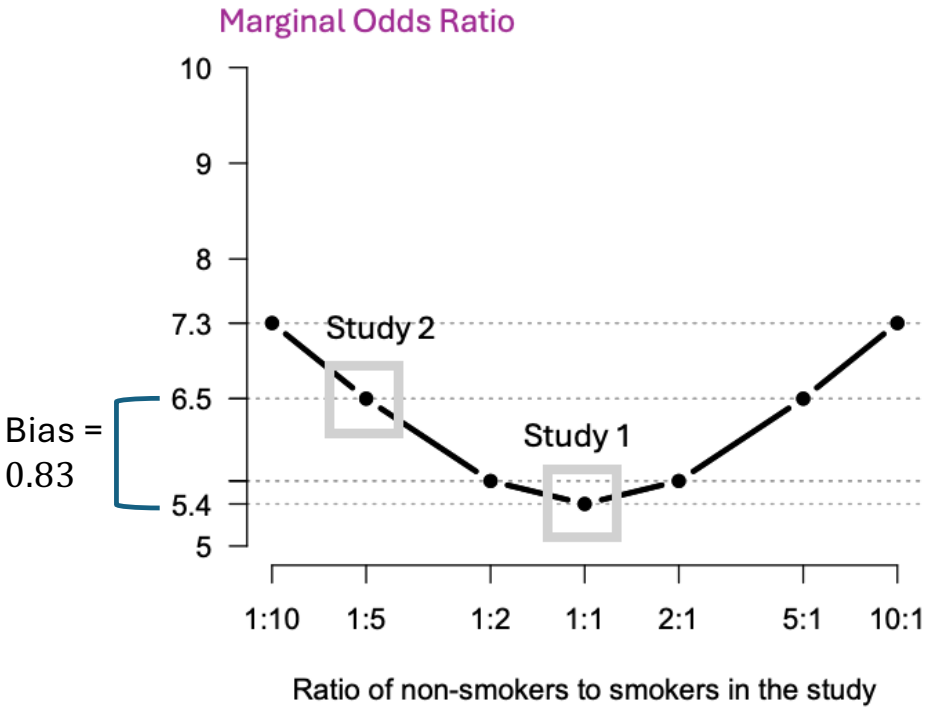
Assumption: *Requires correct specification of the outcome model*

Conclusion: There is no difference between C and B in the Study 2 population

EXAMPLE 2

Marginal odds ratio vs conditional odds ratio

The marginal odds ratio depends on the distribution of prognostic factors in the study population!



Contrary to current recommendations?

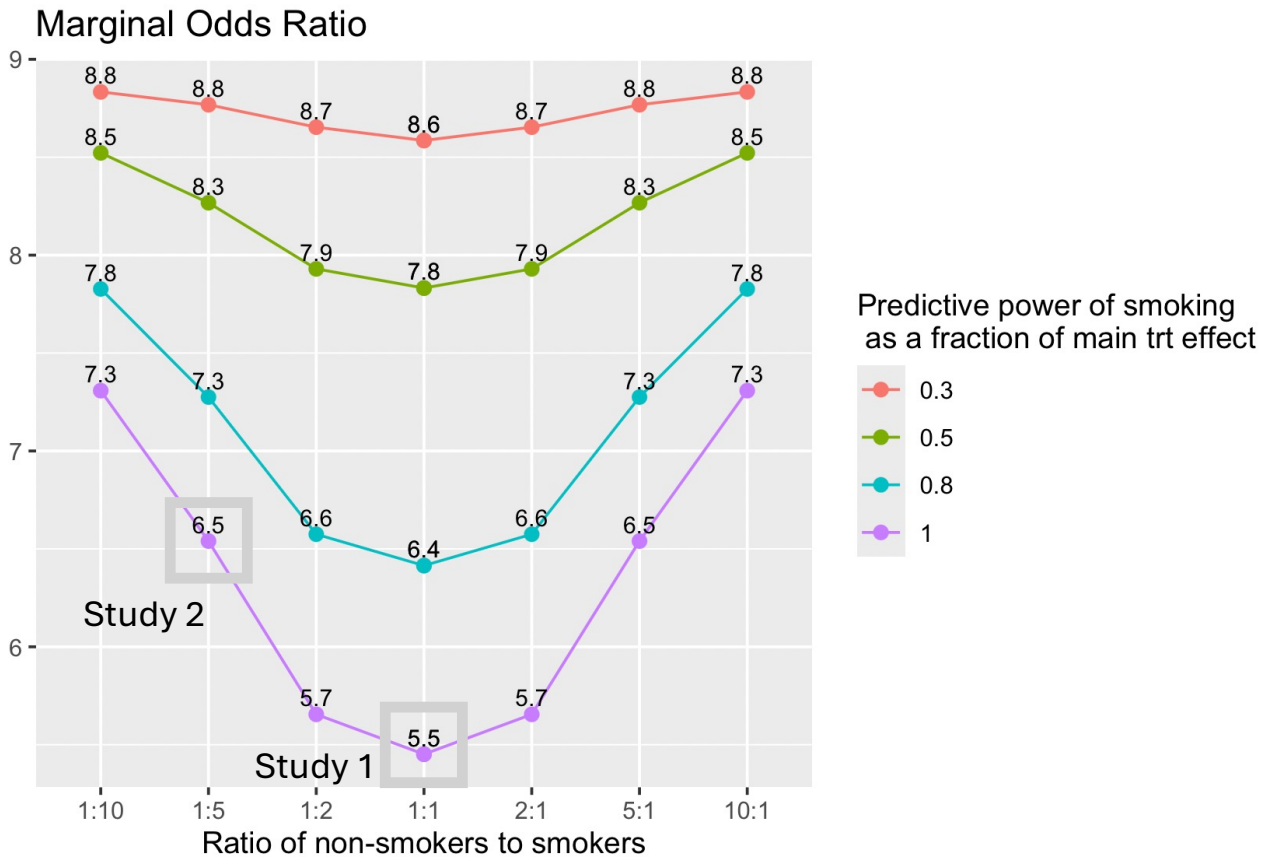
- NICE DSU TSD 18: "To avoid loss of precision due to over-matching, no prognostic variables which are not also effect modifiers should be adjusted for, as variables which are purely prognostic do not affect the estimated relative treatment effect."[1]
- Vo (2023): purely prognostic variables can be "safely excluded."[2]

1 - Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton KJ. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. 2016.

2 - Vo TT. A cautionary note on the use of G-computation in population adjustment. Research Synthesis Methods. 2023; doi:10.1002/jrsm.1621

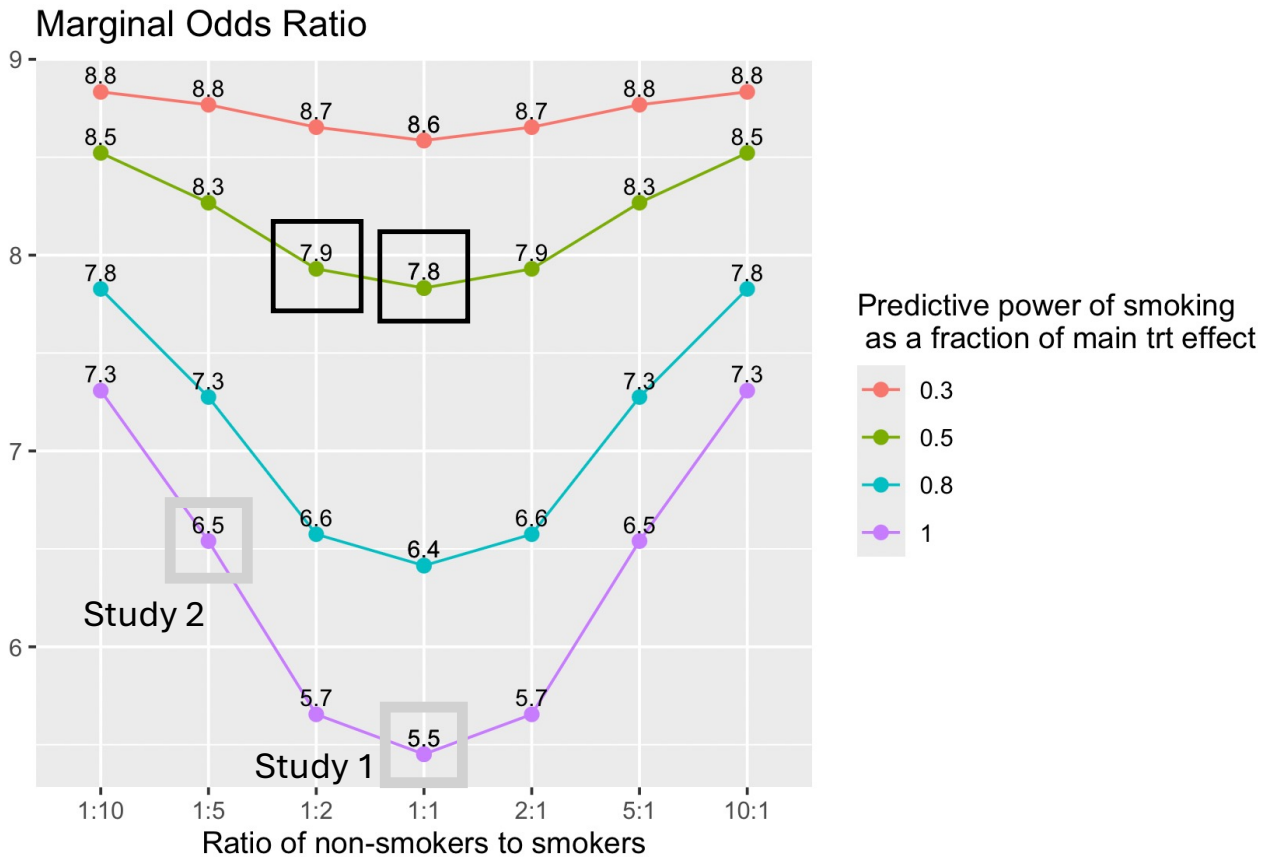
EXAMPLE 2

What if we ignore purely prognostic variables?



EXAMPLE 2

What if we ignore purely prognostic variables?



The bias will only be small unless:

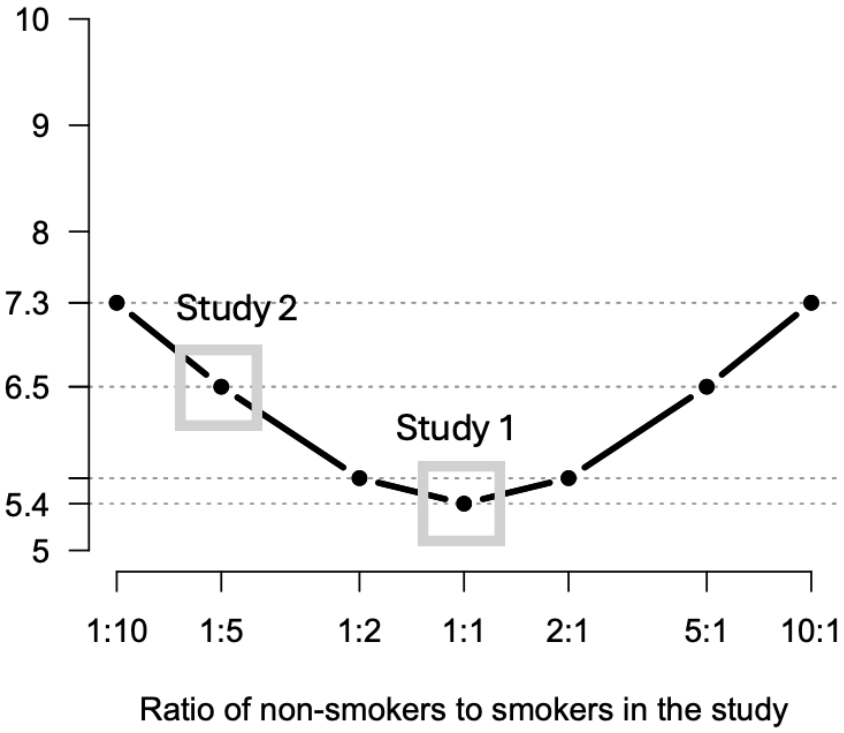
- The ratio of non-smokers to smokers is very different in the two studies
- The predictive effect of smoking is strong

EXAMPLE 2

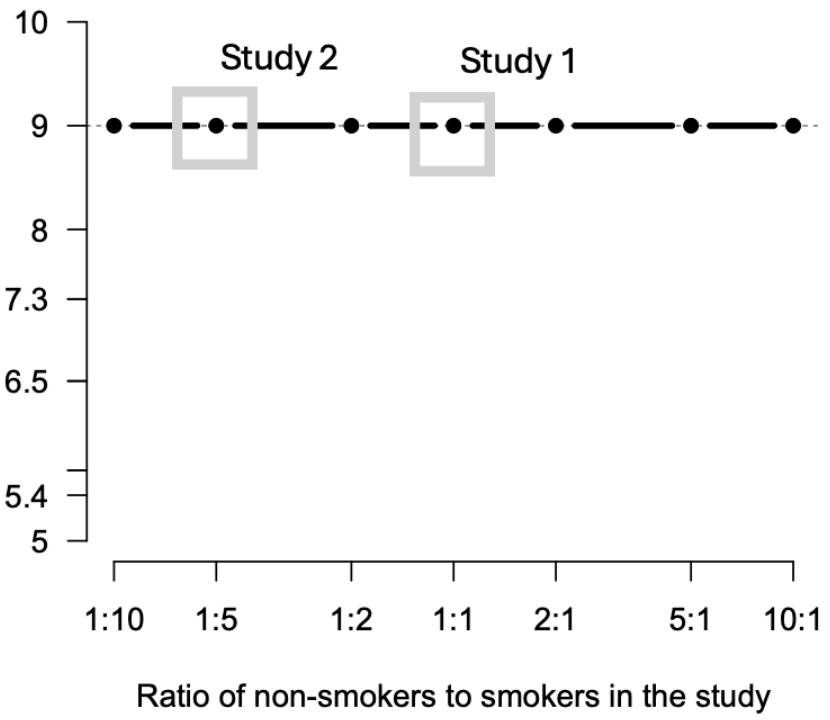
Marginal odds ratio vs conditional odds ratio

The marginal odds ratio depends on the distribution of prognostic factors in the study population!

Marginal odds ratio

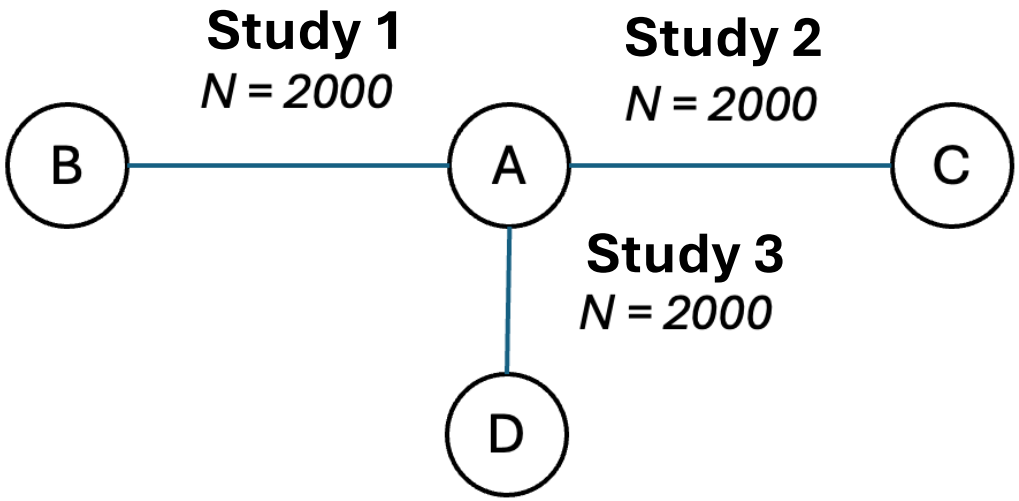


Population-adjusted conditional odds ratio



EXAMPLE 3

Network Meta-Analysis



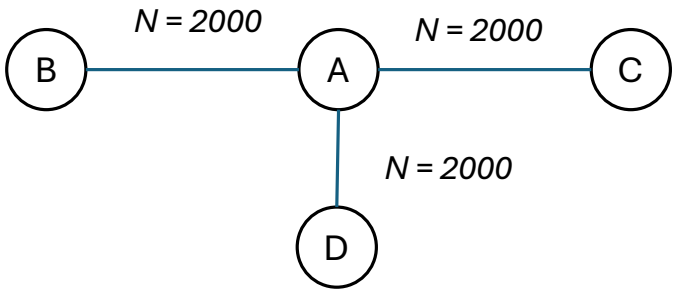
Ratio of non-smokers to smokers is 3:1 in all three studies

Three trials comparing four treatments A, B, C, and D.

EXAMPLE 3

Marginal vs conditional effects

Network Meta-Analysis

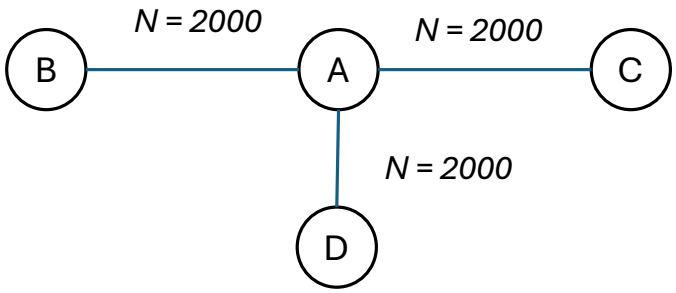


- Outcome of interest: the number of individuals who experience disease progression.
- Smoking status is prognostic ***and*** effect-modifying.

EXAMPLE 3

Marginal vs conditional effects

Network Meta-Analysis



Objective: Rank the treatments according to their effectiveness

Best treatment for who?

We need to think about the causal question of interest.

- Outcome of interest: the number of individuals who experience disease progression.
- Smoking status is prognostic ***and*** effect-modifying.

EXAMPLE 3

Effect modification and non-collapsibility leads to conflicting treatment decisions: a review of marginal and conditional estimands and recommendations for decision-making

David M. Phillippo[1]
Population Health Sciences, Bristol Medical School, University of Bristol, UK

Antonio Remiro-Azócar
Methods and Outreach, Novo Nordisk Pharma, Madrid, Spain

Anna Heath
Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Canada
Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
Department of Statistical Science, University College London, London, United Kingdom

Gianluca Baio
Department of Statistical Science, University College London, London, United Kingdom

Sofia Dias
Centre for Reviews and Dissemination, University of York, York, United Kingdom

A. E. Ades
Population Health Sciences, Bristol Medical School, University of Bristol, UK

Nicky J. Welton
Population Health Sciences, Bristol Medical School, University of Bristol, UK

Phillippo, D. M., Remiro-Azócar, A., Heath, A., Baio, G., Dias, S., Ades, A. E., & Welton, N. J. (2024). Effect modification and non-collapsibility leads to conflicting treatment decisions: a review of marginal and conditional estimands and recommendations for decision-making. *arXiv preprint arXiv:2410.11438*.

EXAMPLE 3

Study 1

Study	Treatment	Non-smokers (N=1500)		Smokers (N=500)		Overall (N=2000)	
		Progression Yes	Progression No	Progression Yes	Progression No	Progression Yes	Progression No
Study 1	A	202	548	156	94	358	642
Study 1	B	89	661	42	208	131	869
Marginal odds ratio: B/A		$\frac{89/661}{202/548} = 0.37$		$\frac{42/208}{156/94} = 0.12$		$\frac{131/869}{358/642} = \mathbf{0.27}$	
Population-average conditional odds ratio: B/A		$\exp(\frac{4500}{6000}\log(0.37) + \frac{1500}{6000}\log(0.12)) = \mathbf{0.28}$					

Based on the marginal odds ratios, treatment B is the best

EXAMPLE 3

Study	Treatment	Non-smokers (N=4500)		Smokers (N=1500)		Overall (N=6000)	
		Progression Yes	Progression No	Progression Yes	Progression No	Progression Yes	Progression No
Study 1	A	202	548	156	94	358	642
Study 1	B	89	661	42	208	131	869
Marginal odds ratio: B/A		$\frac{89/661}{202/548} = 0.37$		$\frac{42/208}{156/94} = 0.12$		$\frac{131/869}{358/642} = \mathbf{0.27}$	
Study 2	A	202	548	156	94	358	642
Study 2	C	13	737	144	106	157	843
Marginal odds ratio: C/A		$\frac{13/737}{202/548} = 0.05$		$\frac{144/106}{156/94} = 0.82$		$\frac{157/843}{358/642} = \mathbf{0.33}$	
Study 3	A	202	548	156	94	358	642
Study 3	D	137	613	6	244	143	857
Marginal odds ratio: D/A		$\frac{137/613}{202/548} = 0.61$		$\frac{6/244}{156/94} = 0.01$		$\frac{143/857}{358/642} = \mathbf{0.30}$	

Based on the conditional odds ratios, treatment B is the worst

EXAMPLE 3

Study	Treatment	Non-smokers (N=4500)		Smokers (N=1500)		Overall (N=6000)	
		Progression Yes	Progression No	Progression Yes	Progression No	Progression Yes	Progression No
Study 1	A	202	548	156	94	358	642
Study 1	B	89	661	42	208	131	869
Marginal odds ratio: B/A		$\frac{89/661}{202/548} = 0.37$		$\frac{42/208}{156/94} = 0.12$		$\frac{131/869}{358/642} = 0.27$	
Pop-avg conditional odds ratio: B/A		$\exp(\frac{4500}{6000}\log(0.37) + \frac{1500}{6000}\log(0.12)) =$ **0.28**					
Study 2	A	202	548	156	94	358	642
Study 2	C	13	737	144	106	157	843
Marginal odds ratio: C/A		$\frac{13/737}{202/548} = 0.05$		$\frac{144/106}{156/94} = 0.82$		$\frac{157/843}{358/642} = 0.33$	
Pop-avg conditional odds ratio: C/A		$\exp(\frac{4500}{6000}\log(0.05) + \frac{1500}{6000}\log(0.82)) =$ **0.10**					
Study 3	A	202	548	156	94	358	642
Study 3	D	137	613	6	244	143	857
Marginal odds ratio: D/A		$\frac{137/613}{202/548} = 0.61$		$\frac{6/244}{156/94} = 0.01$		$\frac{143/857}{358/642} = 0.30$	
Pop-avg conditional odds ratio: D/A		$\exp(\frac{4500}{6000}\log(0.61) + \frac{1500}{6000}\log(0.01)) =$ **0.24**					

Rankings

1. Treatment B
2. Treatment D
3. Treatment C

1. Treatment C
2. Treatment D
3. Treatment B

Marginal and conditional answer two different questions

EXAMPLE 3

- The **marginal effect** results in a decision that minimizes the number of events overall.
- The **population average conditional effect** results in a decision that is optimal for the greatest number of individuals.
- When there is an effect modifier, marginal and conditional effects can lead to **conflicting treatment rankings**.

Also:

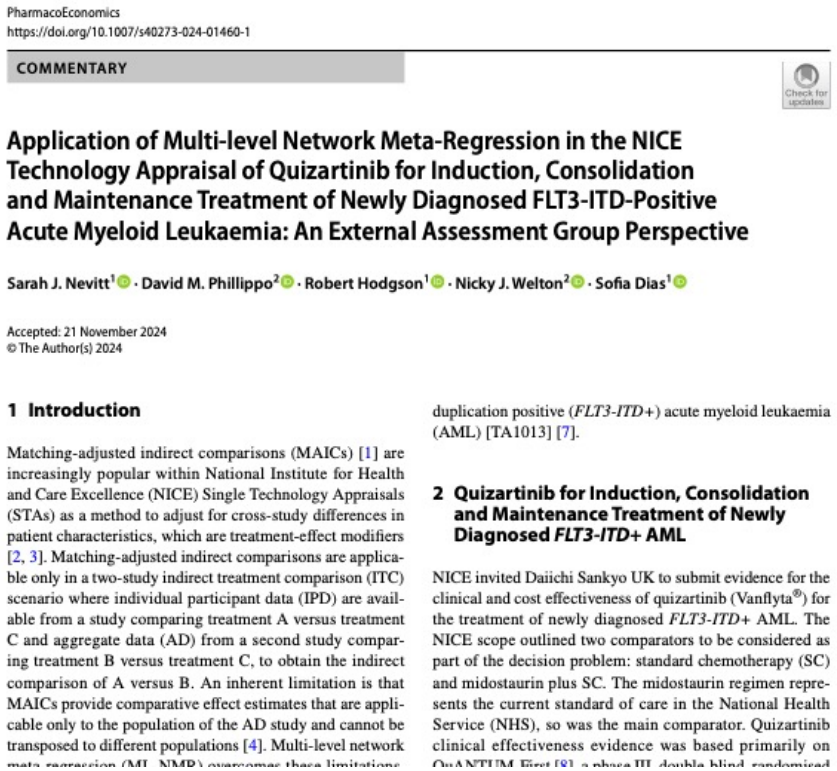
- Conditional effects can be transported to a different target population- not so easy with marginal effects.
- Estimating conditional effects typically requires IPD or additional assumptions.
  - MAIC and STC estimators can only obtain *marginal* effect estimates for the target population defined by the study for which IPD is not available.
  - ML-NMR estimators can obtain conditional effects but require an additional "shared effect-modifier" assumption.

EXAMPLE 3

Recent perspectives

"For decision making, the marginal treatment effect represents the effect of moving everyone within the target population from treatment with midostaurin to treatment with quizartinib."

"Population-average conditional treatment effects are specific to a target population with a given distribution of treatment-effect modifier characteristics and are interpreted as the average of the individual-level treatment effects in the population, i.e. the average effect of moving each individual within the target population from treatment with midostaurin to treatment with quizartinib."



PharmacoEconomics
https://doi.org/10.1007/s40273-024-01460-1

COMMENTARY

Application of Multi-level Network Meta-Regression in the NICE Technology Appraisal of Quizartinib for Induction, Consolidation and Maintenance Treatment of Newly Diagnosed FLT3-ITD-Positive Acute Myeloid Leukaemia: An External Assessment Group Perspective

Sarah J. Nevitt[1] · David M. Phillippo[2] · Robert Hodgson[1] · Nicky J. Welton[2] · Sofia Dias[1]

Accepted: 21 November 2024

1 Introduction

Matching-adjusted indirect comparisons (MAICs) [1] are increasingly popular within National Institute for Health and Care Excellence (NICE) Single Technology Appraisals (STAs) as a method to adjust for cross-study differences in patient characteristics, which are treatment-effect modifiers [2, 3]. Matching-adjusted indirect comparisons are applicable only in a two-study indirect treatment comparison (ITC) scenario where individual participant data (IPD) are available from a study comparing treatment A versus treatment C and aggregate data (AD) from a second study comparing treatment B versus treatment C, to obtain the indirect comparison of A versus B. An inherent limitation is that MAICs provide comparative effect estimates that are applicable only to the population of the AD study and cannot be transposed to different populations [4]. Multi-level network

duplication positive (*FLT3-ITD+*) acute myeloid leukaemia (AML) [TA1013] [7].

2 Quizartinib for Induction, Consolidation and Maintenance Treatment of Newly Diagnosed *FLT3-ITD+* AML

NICE invited Daiichi Sankyo UK to submit evidence for the clinical and cost effectiveness of quizartinib (Vanflyta®) for the treatment of newly diagnosed *FLT3-ITD+* AML. The NICE scope outlined two comparators to be considered as part of the decision problem: standard chemotherapy (SC) and midostaurin plus SC. The midostaurin regimen represents the current standard of care in the National Health Service (NHS), so was the main comparator. Quizartinib clinical effectiveness evidence was based primarily on

Nevitt, S. J., Phillippo, D. M., Hodgson, R., Welton, N. J., & Dias, S. (2025). Application of Multi-level Network Meta-Regression in the NICE Technology Appraisal of Quizartinib for Induction, Consolidation and Maintenance Treatment of Newly Diagnosed FLT3-ITD-Positive Acute Myeloid Leukaemia: An External Assessment Group Perspective. *PharmacoEconomics*, *43*(3), 243-247.

EXAMPLE 3

Recent perspectives

"Senn argues that "we usually treat individuals not population [... but] reimbursement decisions in HTA are typically made on populations not single patients."

"The target population of interest is not necessarily broad. Often, it is highly selected and takes into account specific patient characteristics. For instance, the final scope of a health technology appraisal for abemaciclib (TA810), recently published by the National Institute for Health and Care Excellence, describes the target population as 'adults with hormone receptor-positive, HER2-negative, node-positive early breast cancer after definitive surgery of the primary breast tumor at high-risk of recurrence.'"

Received: 2 August 2022 | Revised: 18 August 2022 | Accepted: 18 August 2022
DOI: 10.1002/sim.9566

COMMENTARY

Statistics in Medicine WILEY

Some considerations on target estimands for health technology assessment

Antonio Remiro-Azócar

Medical Affairs Statistics, Bayer plc, Reading, UK

Correspondence
Antonio Remiro-Azócar, Medical Affairs Statistics, Bayer plc, 400 South Oak Way, Reading RG2 6AD, UK.
Email: antonio.remiro-azocar@bayer.com

First and foremost, I would like to thank an anonymous Associate Editor and Prof. Nigel Stallard for arranging a fascinating discussion around my article, "Target estimands for population-adjusted indirect comparisons."[1] This is based on a prior exchange with Phillippo et al.[2-4] I extend my gratitude to Russek-Cohen,[5] Schiel,[6] Senn,[7] Spieker,[8] and Van Lancker et al,[9] for their additional contributions.

This rejoinder discusses the potential development of an estimands framework in the context of evidence synthesis and health technology assessment (HTA). I consider the following base-case scenario. An evidence synthesis (eg, an indirect treatment comparison or a network meta-analysis) is required for HTA. The evidence synthesis combines the results of multiple randomized controlled trials (RCTs).[10-12] Each RCT has been designed for regulatory approval in the premarketing authorization setting, and has target estimands of its own. There are exceptions to this scenario; regulatory decisions are not exclusively based on RCTs and HTA decisions are not exclusively based on evidence synthesis. Nevertheless, it is

Senn S. Conditions for success and margins of error: estimation in clinical trials. Stat Med. 2022.
Remiro-Azócar, A. (2022). Some considerations on target estimands for health technology assessment. *Statistics in Medicine*, *41*(28), 5592.

What do we need to adjust for?

	EXAMPLE 1 **Unanchored ITC across different populations**	EXAMPLE 2 **Anchored ITC across different populations**	EXAMPLE 3 **Anchored NMA across homogeneous populations**
Risk difference	Prognostics and effect modifiers	Effect modifiers	Nothing
Population-average conditional odds ratio	Prognostics and effect modifiers	Effect modifiers	Effect modifiers
Marginal odds ratio	Prognostics and effect modifiers	Prognostics and effect modifiers*	Nothing

*Failure to adjust for purely prognostic factors will likely result in only minimal bias.

Thank you!

EXAMPLE 2

Study 1

Ratio of non-smokers to smokers is 1:1

Binary outcome (Response)	Non-smokers 		Smokers 		Overall	
	Response Yes	Response No	Response Yes	Response No	Response Yes	Response No
Treatment A	90	10	50	50	140	60
Treatment C	50	50	10	90	60	140
Risk difference: A/C	0.9 – 0.5 = 0.4		0.5 – 0.1 = 0.4		0.7 – 0.3 = 0.4	
Marginal odds ratio: B/C	$\frac{90/10}{50/50} = 9$		$\frac{50/50}{10/90} = 9$		$\frac{140/60}{60/140} = 5.4$	
Conditional odds ratio: B/C	$\frac{90/10}{50/50} = 9$		$\frac{90/10}{50/50} = 9$		$0.5\times9 + 0.5\times9 = 9$	

EXAMPLE 2

Study 2

Ratio of non-smokers to smokers is 5:1

Binary outcome (Response)	Non-smokers 		Smokers 		Overall	
	Response Yes	Response No	Response Yes	Response No	Response Yes	Response No
Treatment B	2250	250	250	250	2500	500
Treatment C	1250	1250	50	450	1300	1700
Risk difference: A/C	0.9 – 0.5 = 0.4		0.5 – 0.1 = 0.4		0.83 – 0.43 = 0.4	
Marginal odds ratio: B/C	$\frac{2250/250}{1250/1250} = 9$		$\frac{250/250}{50/450} = 9$		$\frac{2500/500}{1300/1700} = 6.5$	
Conditional odds ratio: B/C	$\frac{90/10}{50/50} = 9$		$\frac{90/10}{50/50} = 9$		$0.5\times9 + 0.5\times9 = 9$	

EXAMPLE 3

Study 1

Study	Treatment	Non-smokers (N=4500)		Smokers (N=1500)		Overall (N=6000)	
		Progression Yes	Progression No	Progression Yes	Progression No	Progression Yes	Progression No
Study 1	A	202	548	156	94	358	642
Study 1	B	89	661	42	208	131	869

Marginal risk difference: B/A -0.151 -0.456 **-0.227**

Pop avg conditional risk difference: B/A $\frac{4500}{6000}(-0.151) + \frac{1500}{6000}(-0.456) = \mathbf{-0.227}$

Marginal odds ratio: B/A $\frac{89/661}{202/548} = 0.37$ $\frac{42/208}{156/94} = 0.12$ $\frac{131/869}{358/642} = 0.27$

Pop avg conditional odds ratio: B/A $\exp(\frac{4500}{6000}\log(0.37) + \frac{1500}{6000}\log(0.12)) = \mathbf{0.28}$

Study	Treatment	Non-smokers (N=4500)		Smokers (N=1500)		Overall (N=6000)	
		Progression Yes	Progression No	Progression Yes	Progression No	Progression Yes	Progression No
Study 1	A	202	548	156	94	358	642
Study 1	B	89	661	42	208	131	869
Risk difference: B/A		-0.151		-0.456		**-0.227**	
Marginal odds ratio: B/A		$\frac{89/661}{202/548} = 0.37$		$\frac{42/208}{156/94} = 0.12$		$\frac{131/869}{358/642} = 0.27$	
Conditional odds ratio: B/A		$\exp(\frac{4500}{6000}\log(0.37) + \frac{1500}{6000}\log(0.12)) =$ **0.28**					
Study 2	A	202	548	156	94	358	642
Study 2	C	13	737	144	106	157	843
Risk difference: C/A		-0.252		-0.048		**-0.201**	
Marginal odds ratio: C/A		0.05		0.82		0.33	
Conditional odds ratio: C/A		$\exp(\frac{4500}{6000}\log(0.05) + \frac{1500}{6000}\log(0.82)) =$ **0.10**					
Study 3	A	202	548	156	94	358	642
Study 3	D	137	613	6	244	143	857
Risk difference: D/A		-0.087		-0.600		**-0.215**	
Marginal odds ratio: D/A		0.61		0.01		0.30	
Conditional odds ratio: D/A		$\exp(\frac{4500}{6000}\log(0.61) + \frac{1500}{6000}\log(0.01)) =$ **0.24**					

Rankings

1. Treatment B
2. Treatment D
3. Treatment C

1. Treatment C
2. Treatment D
3. Treatment B

1. Treatment B
2. Treatment D
3. Treatment C