Design and Methods for Comparative Non-randomized Studies

Symposium "Registry-based non-randomized studies for treatment comparisons“
26 May 2026, Göttingen

Uwe Siebert

Professor of Public Health, Medical Decision Making and Health Technology Assessment (UMIT TIROL)
Adjunct Professor of Epidemiology and Health Policy & Management (Harvard Chan School of Public Health)
Institute Head at UMIT TIROL – University for Health Sciences and Technology, Austria

Uwe Siebert

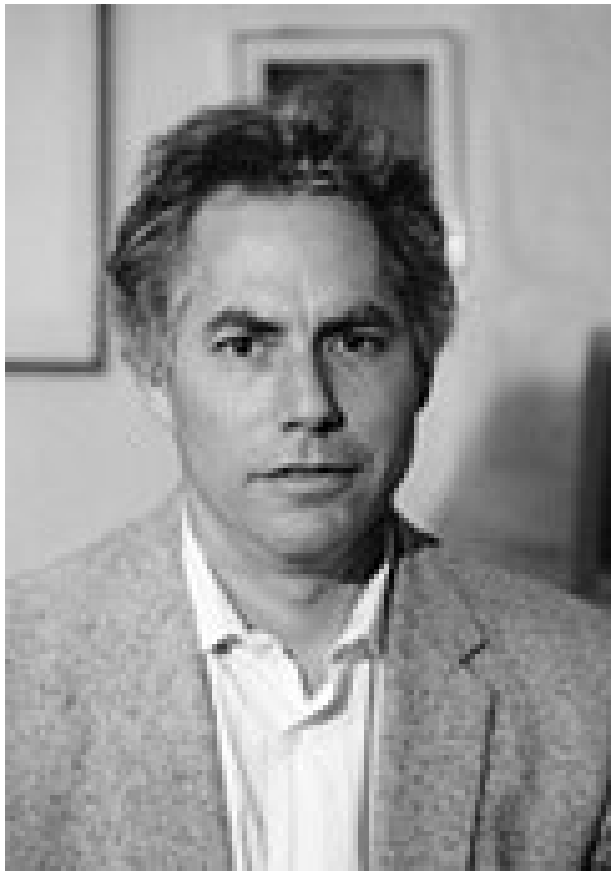
- Head & Professor, Institute of Public Health, Medical Decision Making and Health Technology Assessment, UMIT TIROL - University for Health Sciences and Technology, Hall in Tirol, Austria
- Adjunct Professor of Epidemiology and Health Policy & Management, Harvard Chan School of Public Health, Boston
- Cardiovascular Research Program, Center for Health Technology Assessment, Mass General Brigham, Harvard Medical School, Boston
- Director, HTADS Continuing Education Program (www.htads.org)
- President ISPOR, Former President SMDM
- **Follow:** @UweSiebert9, LinkedIn: uwe-siebert9, www.htads.org
- **Contact:** uwe.siebert@umit-tirol.at

Drop email with questions after talk

Scan!

- **Research interests:**
  - Study design and causal inference analysis of large real-world observational data and pragmatic trials
  - Linking causal inference, machine learning, decision-analytic modeling and AI to support decision making
- **Teaching**
  - Health Data & Decision Analysis, HTA, Epidemiology, Causal "Big Data" Analysis, Machine Learning/Causal AI

I am honored to dedicate this keynote to my mentor Jamie Robins, whose pioneering insights solved the problem of time-varying confounding 40 years ago and laid the foundation for much of modern causal inference.



James M. Robins

Joint with the Isaac Newton Institute (INI), we are excited to welcome Prof. **James Robins** from Harvard University, to present **"Forty years of causal inference: Report of a great-grandfather"** in the Rothschild Public Lecture, Tuesday, May 26, 2026

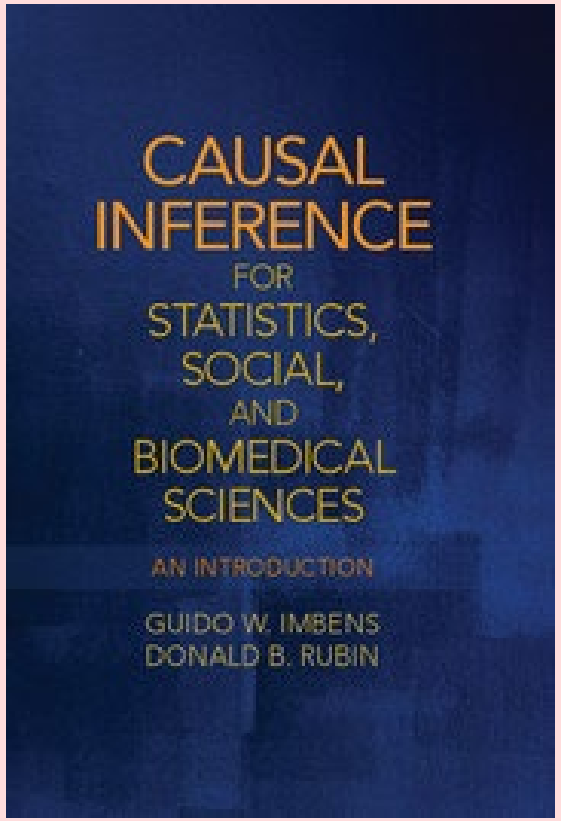
"Robins cut the Gordian knot by inventing a statistic called the g-estimator that makes analysis of data that are simultaneously confounders and intermediate steps possible. […] After a long period of seeking converts to his unconventional methods, Professor James Robins is now considered to be one of the leading mathematical statisticians in the world."

Harvard Public Health Review, Summer 2002:42-43

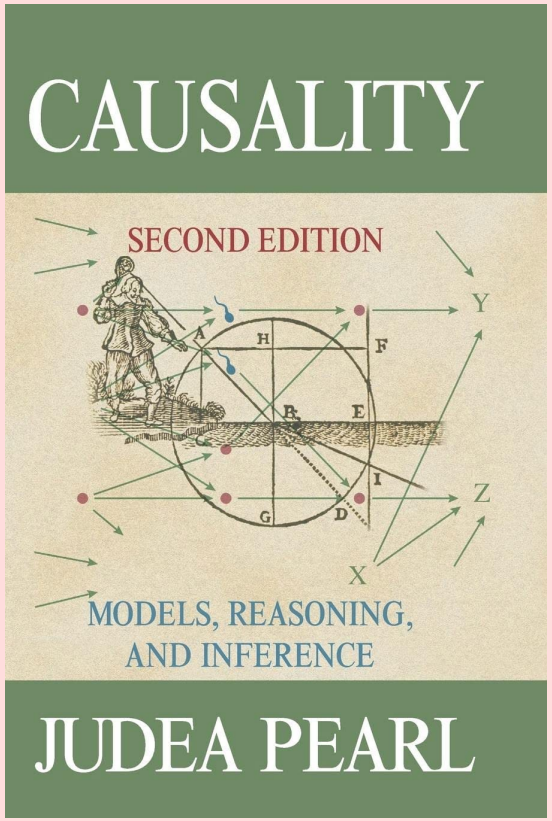
*Literature

- **Greenland S, Pearl J, Robins JM.** Causal diagrams for epidemiologic research. Epidemiology (Cambridge, Mass) 1999; 10: 37-48. [Causal diagrams].
- **Robins JM, Hernán MA, Siebert U.** Estimations of the effects of multiple interventions. In: Ezzati M, Lopez A, Rodgers A, Murray C, eds, Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization. 2004; 2191-230. Internet: http://www.who.int/publications/cra/chapters/volume2/2191-2230.pdf?ua=1 [Methods overview and first application of parametric g-formula]
- **Hernan MA, Robins JM**. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. American Journal of Epidemiology 2016; 183: 758-64. [Target trial, Big data].
- **Kuehne F, Jahn B, Conrads-Frank A, Bundo M, Arvandi M, Endel F, Popper N, Endel G, Urach C, Gyimesi M, Murray E, Danaei G, GazianoTA, Pandya A, Siebert U**. Guidance for a causal comparative effectiveness analysis emulating a target trial based on big real world evidence: when to start statin treatment. J Comp Eff Res 2019;8(12):1013-25. [Target trial protocol, Example].
- **Kuehne F, Arvandi M, Hess LM, Faries DE, Matteucci Gothe R, Gothe H, Beyrer J, Zeimet AG, Stojkov I, Mühlberger N, Oberaigner W, Marth C, Siebert U.** Causal analyses with target trial emulation for real-world evidence removed large self-inflicted biases: systematic bias assessment of ovarian cancer treatment effectiveness. J Clin Epidemiol. 2022 Dec;152:269-280. doi: 10.1016/j.jclinepi.2022.10.005. [Target trial analysis, Example Oncology].
- **Hernan MA, Robins JM.** Per-Protocol Analyses of Pragmatic Trials. NEJM 2017; 377: 1391-8 [IPCW, Pragmatic trials].
- **Kühne F, Schomaker M, Stojkov I, Jahn B, Conrads-Frank A, Siebert S, Sroczynski G, Puntscher S, Schmid D, Schnell-Inderst P, Siebert U.** Causal evidence in health decision making: methodological approaches of causal inference and health decision science. Ger Med Sci. 2022;20:Doc12, 1-24. doi: 10.3205/000314. [Causal inference for decision modeling].

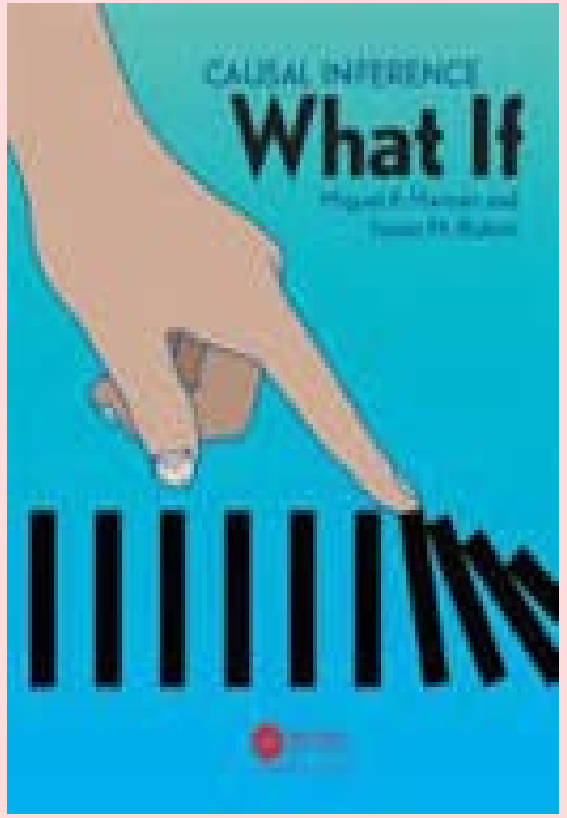
*Textbooks



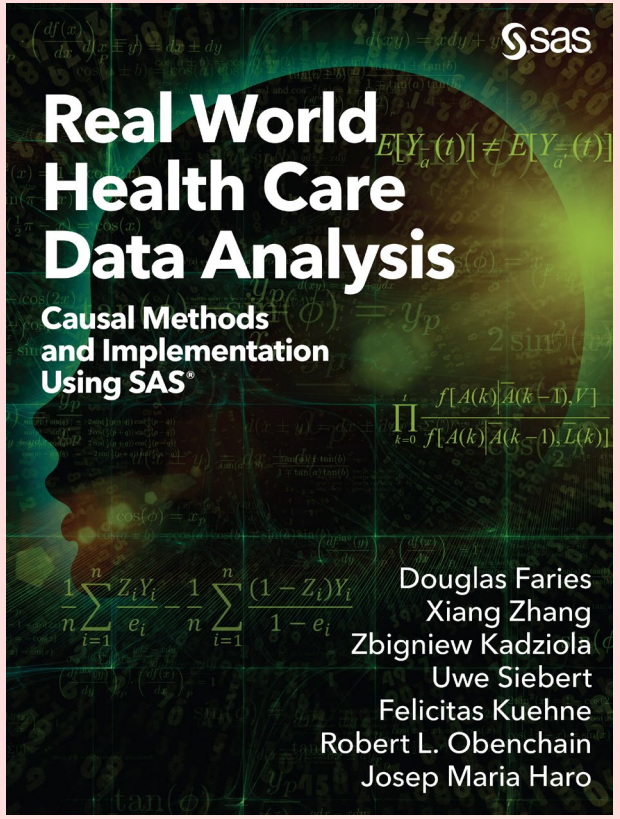
Causal Inference
(Economics)



Causal Inference
(Computer Science)

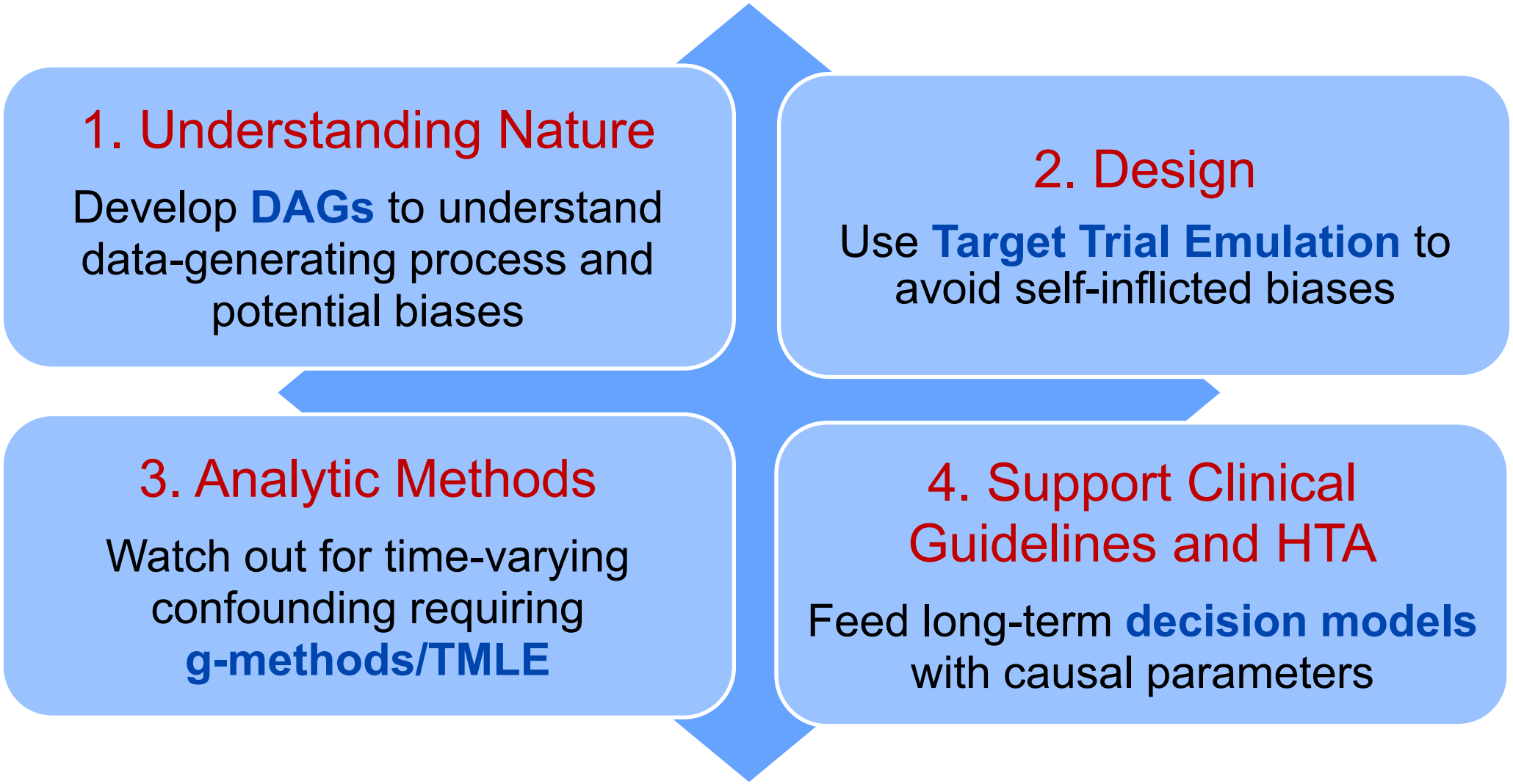


Causal Inference
(Epidemiology)

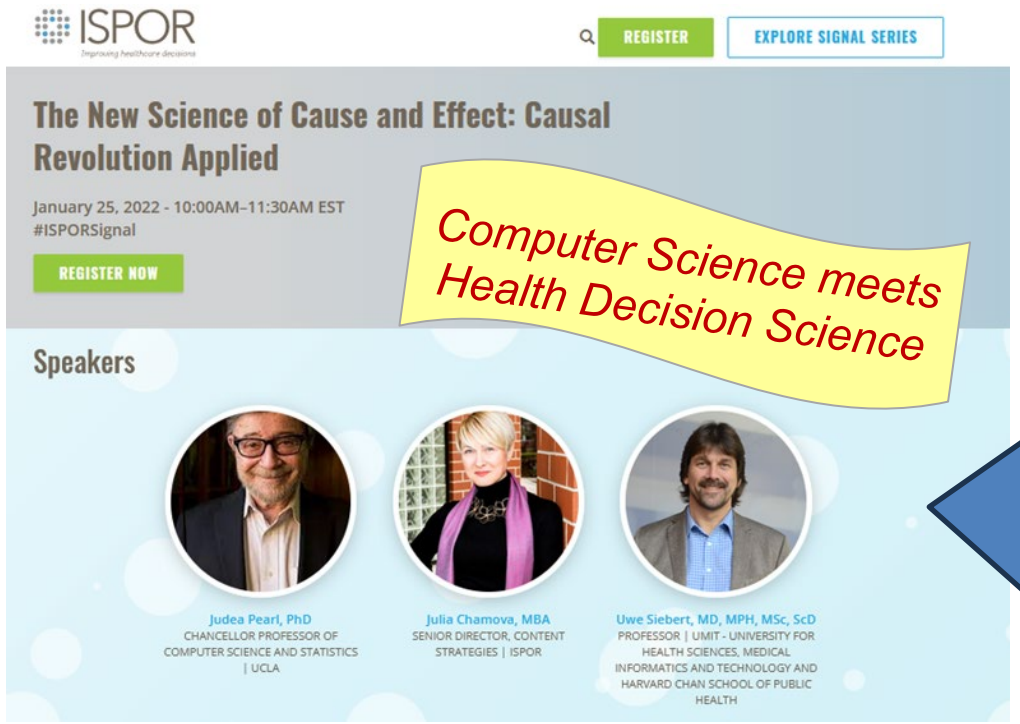


Causal Inference
(RWD Analysis, SAS)

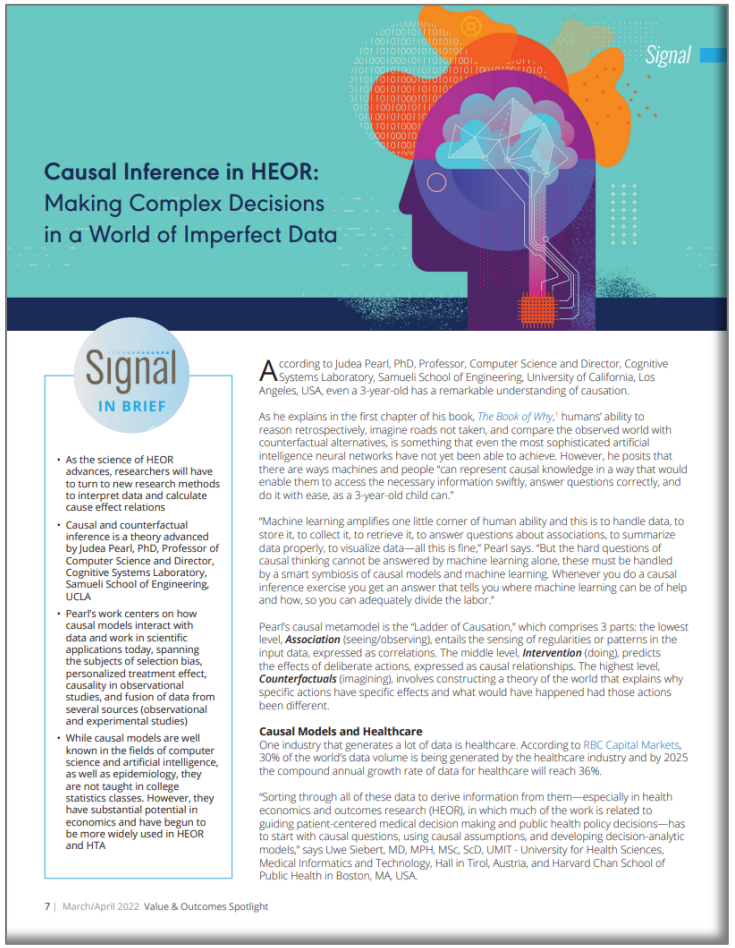
4 Key Elements of a Causal Health Decision Framework



ISPOR SIGNAL and Value & Outcomes Spotlight



" … one of the main tasks in applying causal inference methods in health economics, outcomes research and HTA is understanding which analytical method works best for which type of research question, and recommending what additional evidence should be generated."



See the related Value & Outcomes Spotlight article:
https://www.ispor.org/publications/journals/value-outcomes-spotlight
https://www.ispor.org/docs/default-source/default-document-library/ispor_vos_april-2022_online.pdf?sfvrsn=3264b13_0

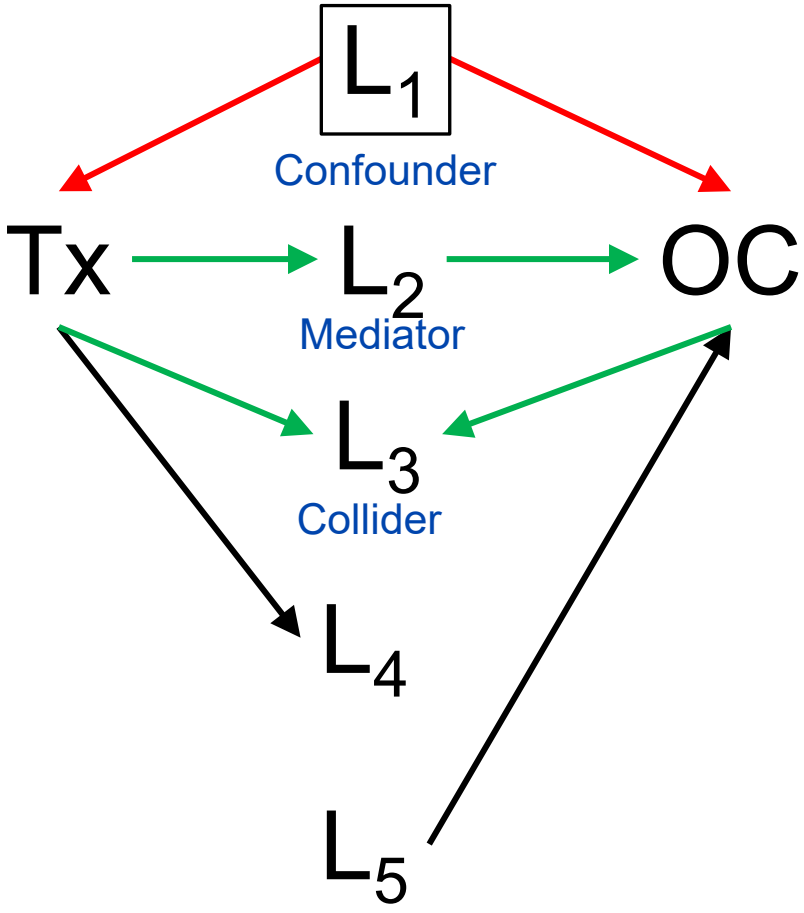
UMIT TIROL

Causal Graphs

Causal Diagrams
(Directed Acyclic Graphs, DAGs)

Tx: Treatment
OC: Outcome
L: Covariate

- Graph is directed (arrows) and acyclic (no loops)
- Causal DAG: assumptions = absence of arrows
- 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:

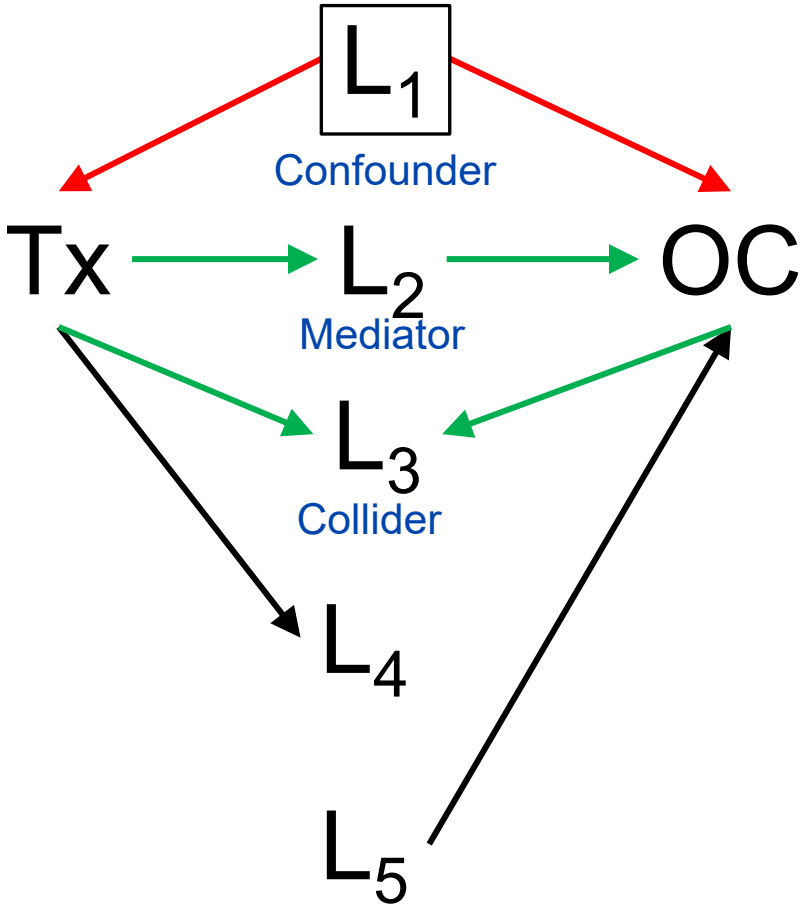
Open backdoor path

Adjustment:

Block (control/adjust for) all open backdoor paths

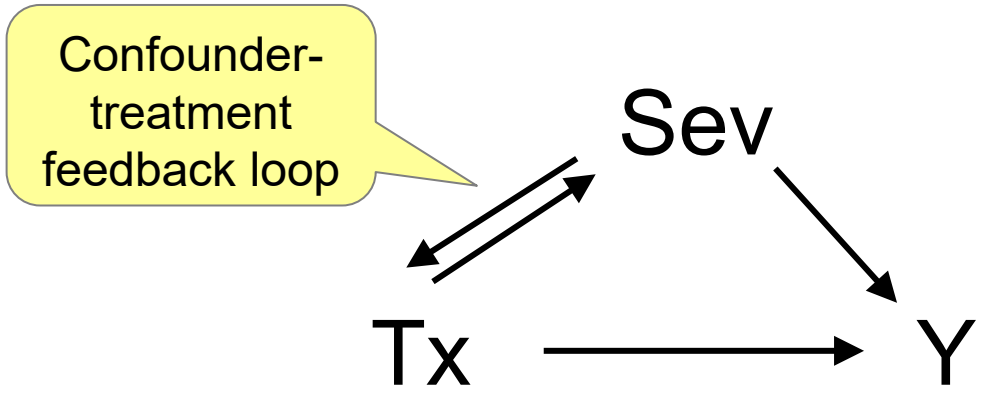
Rule:

➡ *Never control for the future of the treatment*



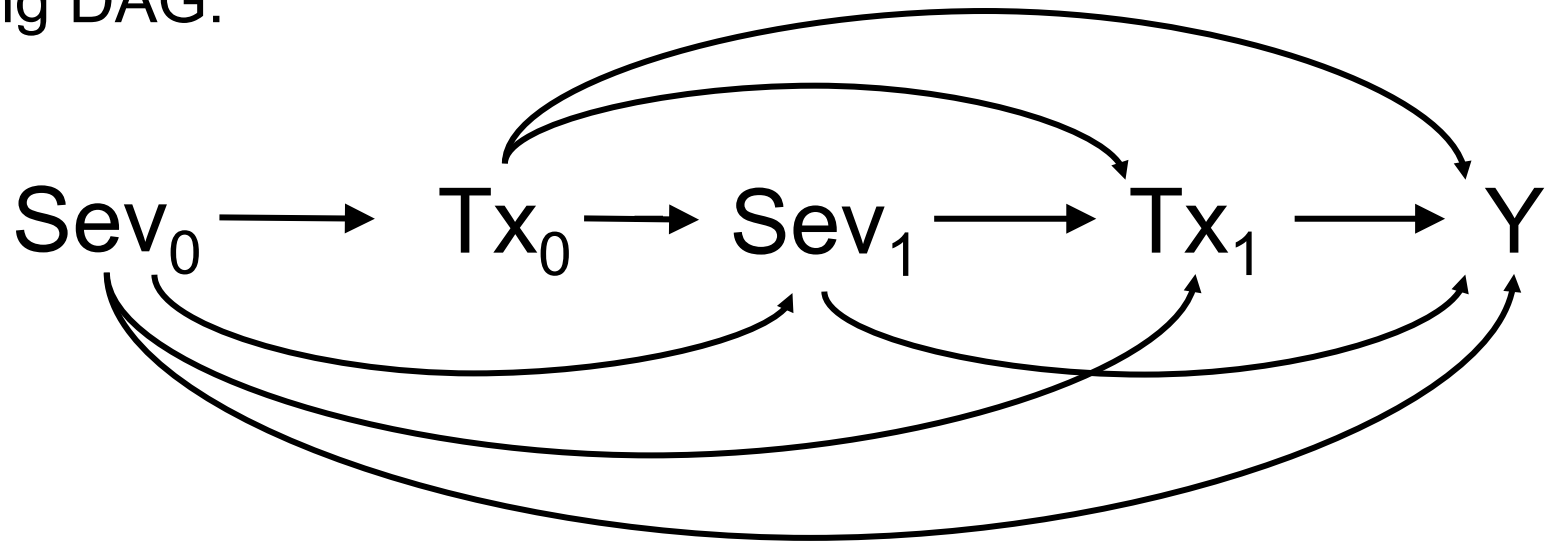
DAGs for Time-dependent Confounding

No DAG:

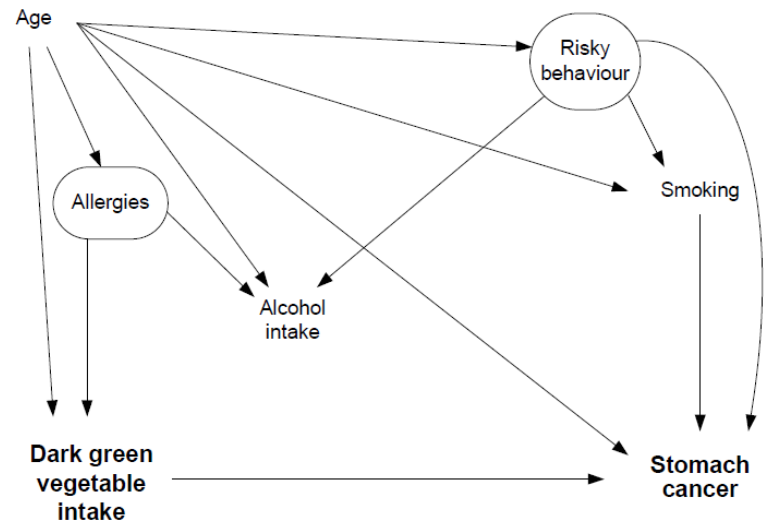


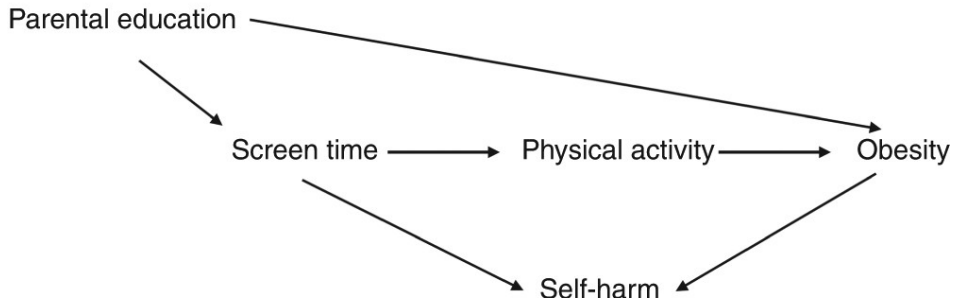
Tx = Treatment
Sev = Severity

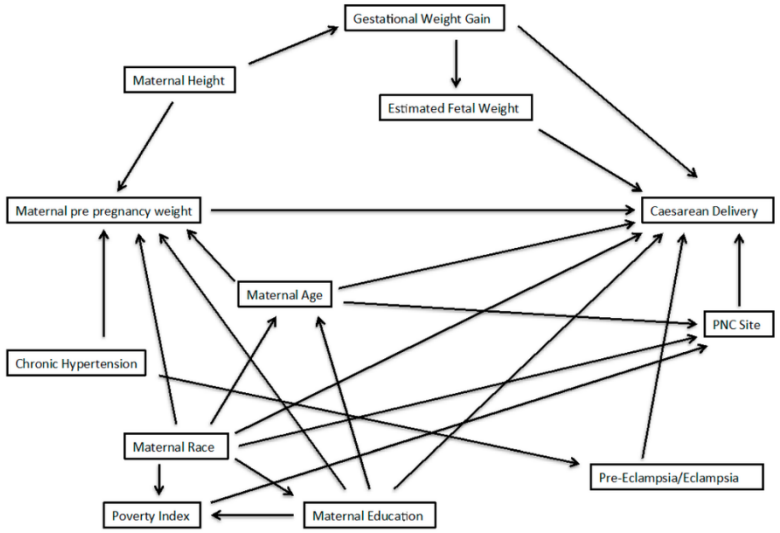
Time-varying DAG:

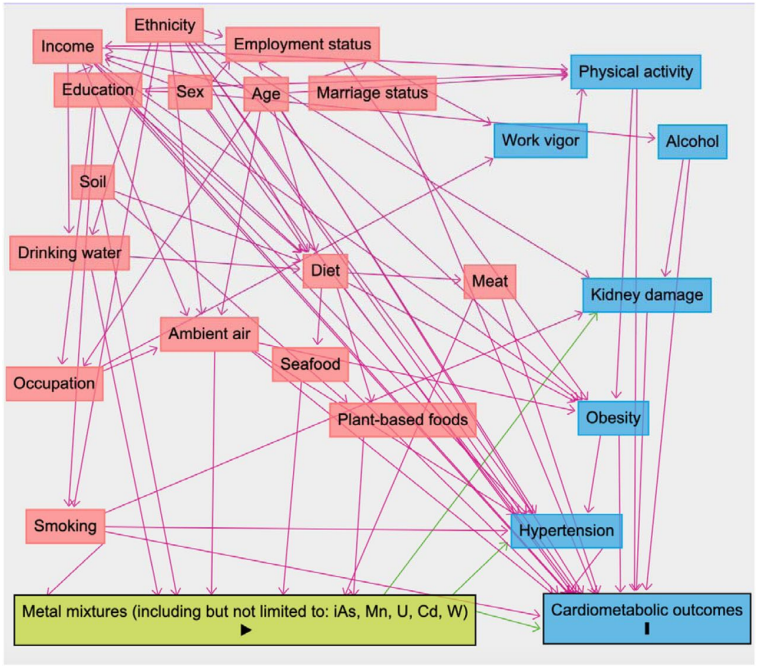


DAG Examples from the Literature



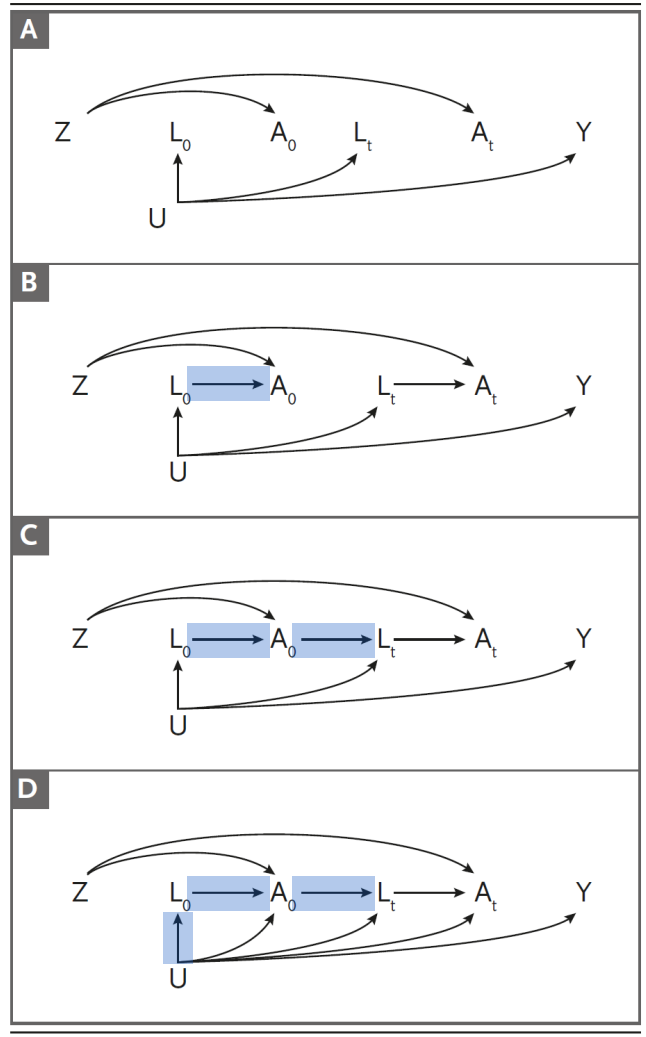






Source: Internet

*Possible DAGs in Pragmatic Trials



No Confounding

Simple Confounding

Time-dependent Confounding

Unmeasured Confounding

Hernán & Robins, NEJM 2017

Design: Target Trial Emulation

(for deeper discussion see presentation of Vanessa Didelez)

Target Trial Approach

American Journal of Epidemiology

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 Emulation to Avoid Self-Inflicted Biases

JAMA Network | Open

Original Investigation | Statistics and Research Methods

Reporting of Observational Studies Explicitly Aiming to Emulate Randomized Trials

A Systematic Review

Harrison J. Hansford, BSc(Hons); Aidan G. Cashin, PhD; Matthew D. Jones, PhD; Sonja A. Swanson, ScD; Nazrul Islam, MD, PhD; Susan R. G. Douglas, BExPhys; Rodrigo R. N. Rizzo, PhD; Jack J. Devonshire, BSc(Hons); Sam A. Williams, BSc(Hons); Issa J. Dahabreh, MD, ScD; Barbra A. Dickerman, PhD; Matthias Egger, MD, PhD; Xabier Garcia-Albeniz, MD, PhD; Robert M. Golub, MD; Sara Lodi, PhD; Margarita Moreno-Betancur, PhD; Sallie-Anne Pearson, PhD; Sebastian Schneeweiss, MD, ScD; Jonathan A. C. Sterne, PhD; Melissa K. Sharp, PhD; Elizabeth A. Stuart, PhD; Miguel A. Hernán, MD, DrPh; Hopin Lee, PhD; James H. McAuley, PhD

Introduction

Analyses of observational (nonexperimental) data can be used to estimate the causal effect of interventions when randomized clinical trials are unavailable or infeasible. Bias in observational analyses may be limited by conceptualizing them as attempts to emulate target trials, ie, hypothetical randomized trials that would answer causal questions of interest.[1-3] Hernán and Robins[4] have

REFERENCES

1. Kuehne F, Arvandi M, Hess LM, et al. Causal analyses with target trial emulation for real-world evidence removed large self-inflicted biases: systematic bias assessment of ovarian cancer treatment effectiveness. *J Clin Epidemiol*. 2022;152:269-280. doi:10.1016/j.jclinepi.2022.10.005

2. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat Med*. 2019;25(10):1601-1606. doi:10.1038/s41591-019-0597-x

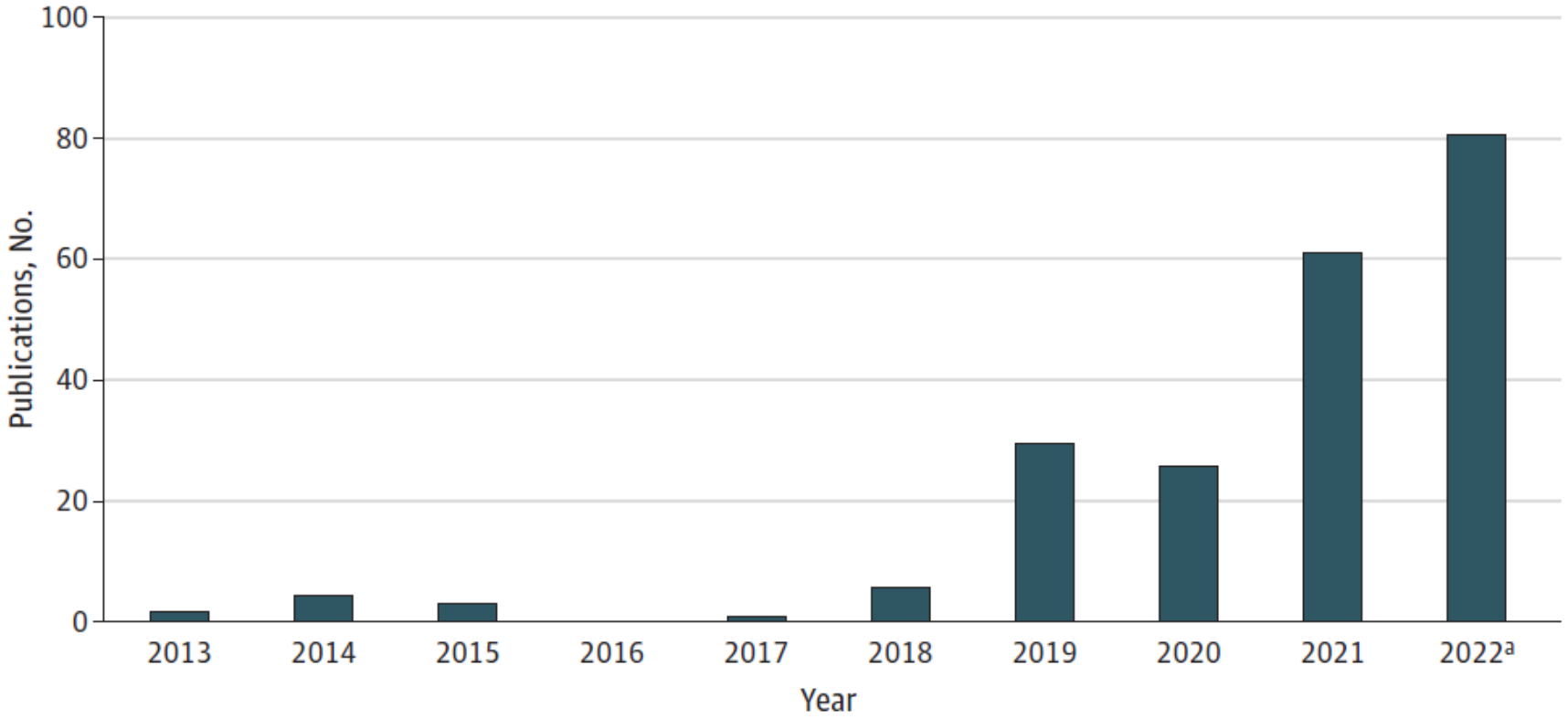
3. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70-75. doi:10.1016/j.jclinepi.2016.04.014

4. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254

https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2809945

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

Target Trial Specification and Emulation

TrAnsparent ReportinG of observational studies Emulating a Target trial (TARGET) guideline

Research Question

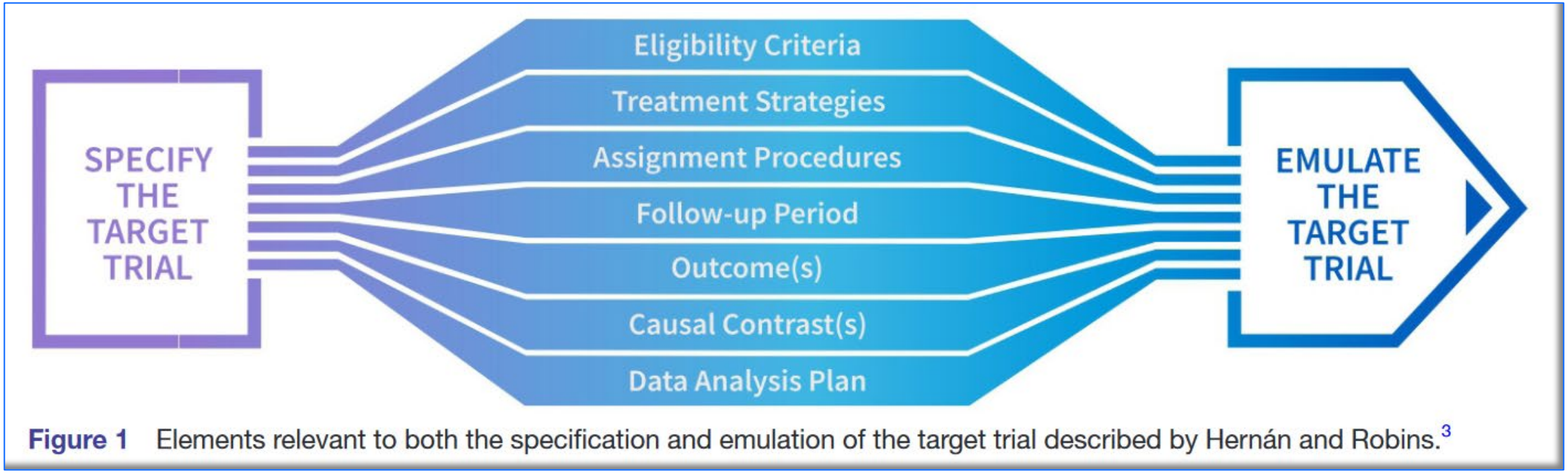


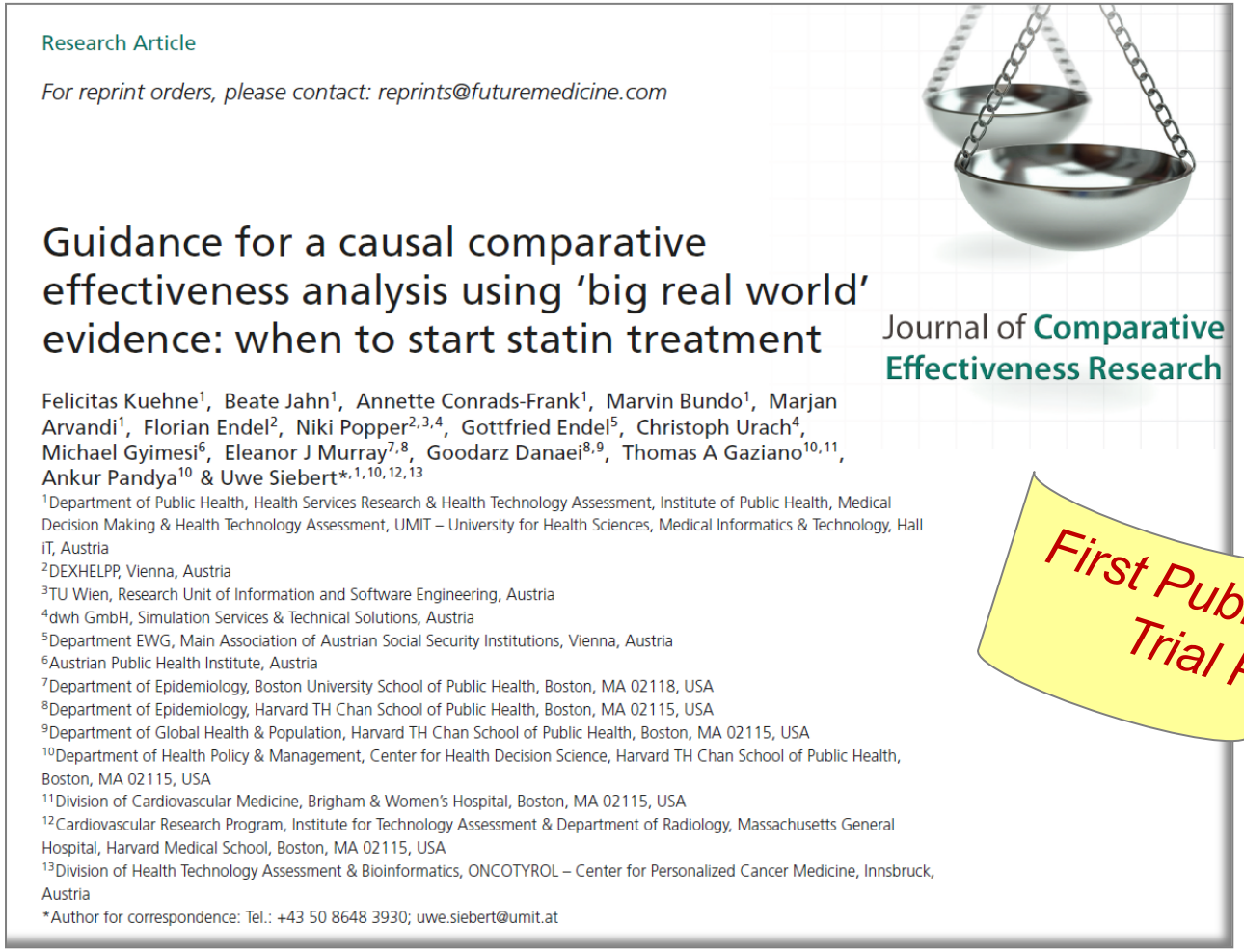
Figure 1 Elements relevant to both the specification and emulation of the target trial described by Hernán and Robins.[3]

Hansford et al, BMJ Open 2023

[3] Hernán & Robins, Am J Epi 2014

UMIT TIROL

Write and Publish Your Target Trial Protocol



Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Guidance for a causal comparative effectiveness analysis using 'big real world' evidence: when to start statin treatment

Felicitas Kuehne[1], Beate Jahn[1], Annette Conrads-Frank[1], Marvin Bundo[1], Marjan Arvandi[1], Florian Endel[2], Niki Popper[2,3,4], Gottfried Endel[5], Christoph Urach[4], Michael Gyimesi[6], Eleanor J Murray[7,8], Goodarz Danaei[8,9], Thomas A Gaziano[10,11], Ankur Pandya[10] & Uwe Siebert*[,1,10,12,13]

[1]Department of Public Health, Health Services Research & Health Technology Assessment, Institute of Public Health, Medical Decision Making & Health Technology Assessment, UMIT – University for Health Sciences, Medical Informatics & Technology, Hall iT, Austria

[2]DEXHELPP, Vienna, Austria

[3]TU Wien, Research Unit of Information and Software Engineering, Austria

[4]dwh GmbH, Simulation Services & Technical Solutions, Austria

[5]Department EWG, Main Association of Austrian Social Security Institutions, Vienna, Austria

[6]Austrian Public Health Institute, Austria

[7]Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118, USA

[8]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA 02115, USA

[9]Department of Global Health & Population, Harvard TH Chan School of Public Health, Boston, MA 02115, USA

[10]Department of Health Policy & Management, Center for Health Decision Science, Harvard TH Chan School of Public Health, Boston, MA 02115, USA

[11]Division of Cardiovascular Medicine, Brigham & Women's Hospital, Boston, MA 02115, USA

[12]Cardiovascular Research Program, Institute for Technology Assessment & Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA

[13]Division of Health Technology Assessment & Bioinformatics, ONCOTYROL – Center for Personalized Cancer Medicine, Innsbruck, Austria

*Author for correspondence: Tel.: +43 50 8648 3930; uwe.siebert@umit.at

Journal of Comparative Effectiveness Research

First Published Target Trial Protocol

Kuehne et al., JCER 2019

Target Trial Protocols for Using RWE

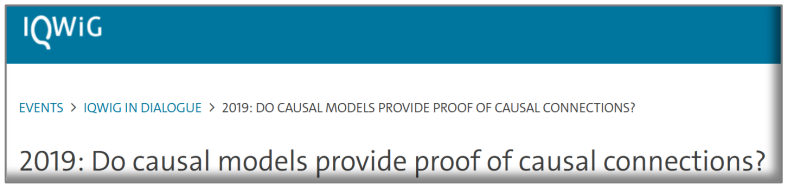
- TTE is often applied to preexisting observational RWD
- In these cases, the fully specified target trial protocol may need to be iteratively adapted as investigators learn about the data (e.g. missing data in confounder variables, coding of outcomes)
- The protocol should include clear rules governing the adaptive process and the investigators' decisions to adapt the protocol to the data, in order to maintain trust in the resulting RWE

Hernán MA, et al., Epidemiology 2026 May 1

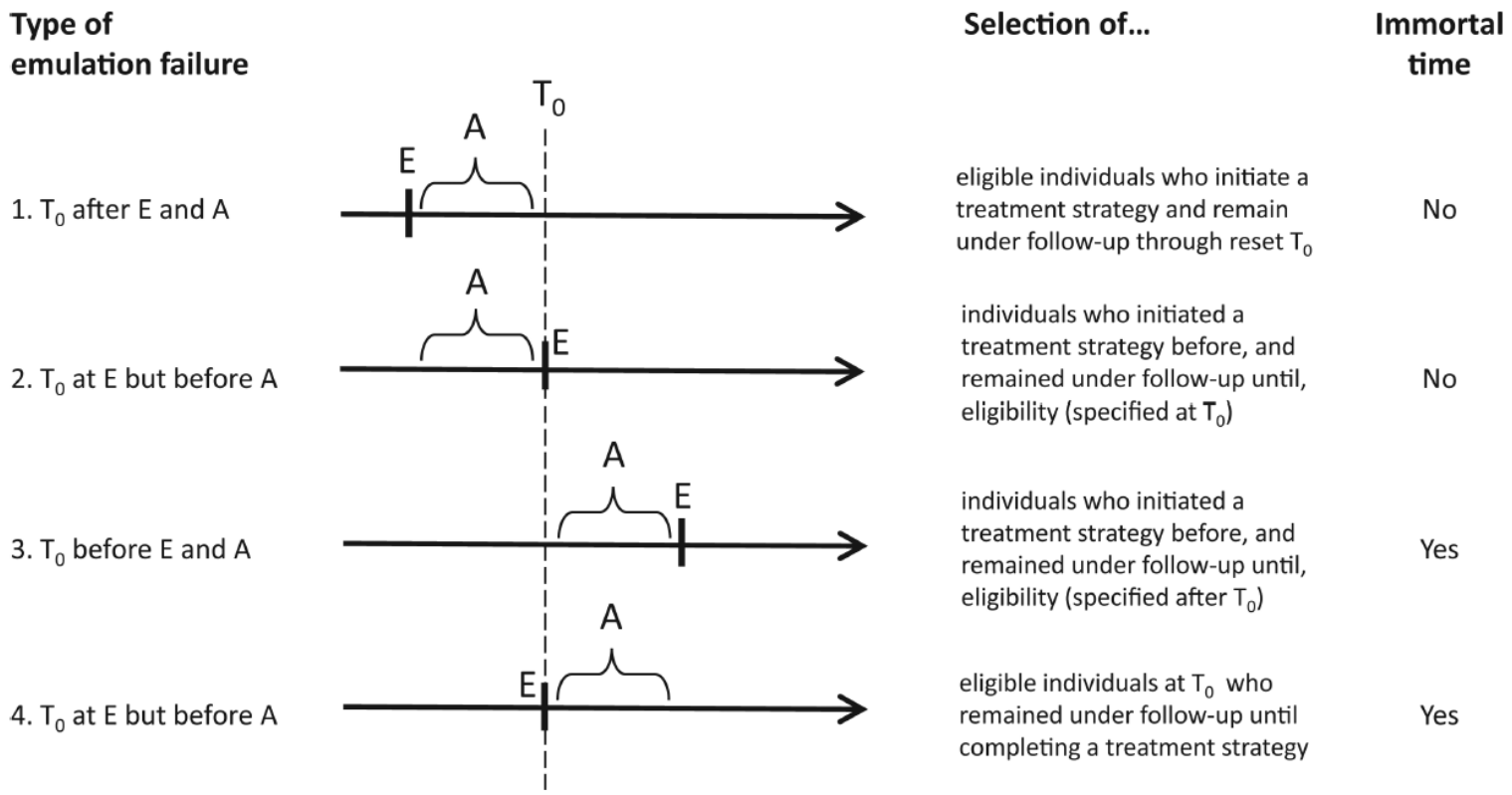
Target Trial Emulation in EU Agencies

- NICE (UK)
  - Describes target trial emulation in methods guidance (June 2022) www.nice.org.uk/corporate/ecd9

- IQWiG (Germany)
  - Discussed causal topics in public events https://www.iqwig.de/en/events/iqwig-in-dialogue/
  - Recently open for causal modeling of screening https://www.iqwig.de/en/presse/press-releases/press-releases-detailpage_75008.html
  - Explicit framework of target trial approach?



Self-inflicted Time-Related Biases ("Misalignment")



Immortal Time due to Selection Bias

Immortal Time due to Information Bias

Fig. 1.
Four examples of failures of emulation of a target trial using observational data. T_0, time zero; E, eligibility; A, period during which treatment strategies are assigned.

Hernán et al., J Clin Epi 2016

Different Approaches to TTE

1) Unique and clear treatment start → Traditional approach
   - All individuals meet eligibility only once; treatment strategies distinguishable at time of eligibility
2) Repeat eligibility → Sequential trials approach
   - Some individuals repeatedly eligible during multiple times and can be assigned to a treatment strategy at multiple times (multiple times zero)
3) Complying with multiple Tx strategies → Clone-censor-weight approach
   - Some individuals belong to multiple treatment arms during some times, especially at baseline ("indistinguishable treatment strategies")

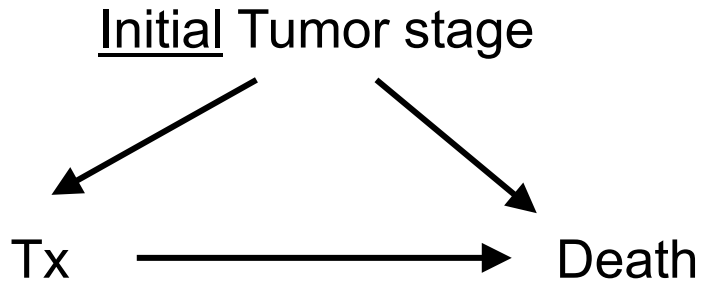
Causal Inference in Indirect Treatment

- 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

Analytic Methods for Causal Inference

Motivation: Confounding

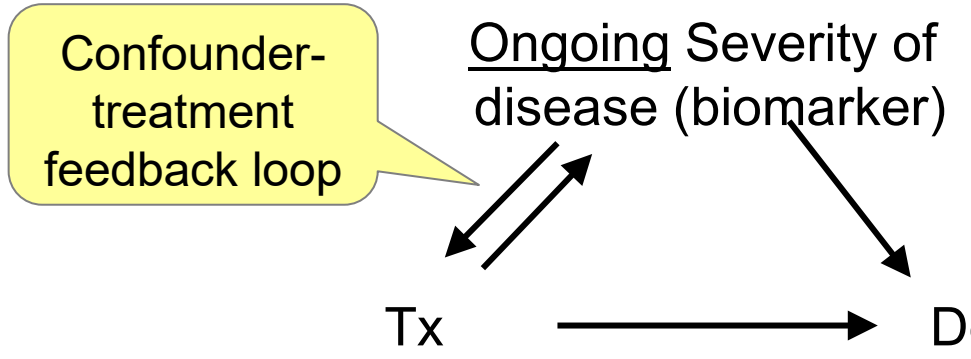
Tx = Treatment



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]

- Regression
- Stratification
- Matching
- IPW

Time-dependent (Post-baseline) Confounding

- g-Methods
  - g-Formula
  - g-Estimation
  - Inverse probability weighting
- Further approaches:
  - Doubly robust methods (TMLE)

- g-Methods
  - g-formula
  - g-estimation
  - inverse probability weighting

Principles of Causal Methods

(g-methods: Robins 1986-1998)

Necessary to control for time-varying confounding

- g-Formula / g-Computation
  - Stratify (standardize) for history, move through time intervals
- g-Estimation with structural nested models (SNM)
  - Calculate counterfactual outcomes, correlate counterfactuals with treatment conditional on confounders
- Inverse probability weighting (IPW) with marginal structural models (MSM)
  - Each individual undergoes each treatment → "clones", weighting; yields unconfounded pseudo population dataset
- Doubly robust methods (TMLE)
  - Simultaneously correct for the confounder-treatment and the confounder outcome relation, leading to bias elimination, if one model is correctly specified

*History

- g-estimation (Robins, 1986) and marginal structural models (Robins, 1998)
  - for situations with time-varying confounding
- Causal network theory (Pearl, late 1990s)
  - for causal inferences, confounder identification must be based on a causal network linking the variables under study
- Causal diagrams (Greenland, Pearl, and Robins, 1999)
  - valuable and visually understandable tool
- First application of g-formula (2002: ScD thesis Siebert with Robins, Hernán)
- Acceptance of methods at regulatory agencies (2009: TA 179)
  - NICE appraisals accepted causal methods to adjust for switching in clinical trial and drug was reimbursed in UK
- Software (last years)
  - some macros available and under development

Machine Learning in Causal Inference

Machine learning (ML) methods have

- the potential to smooth over the data without overly restrictive assumptions regarding the functional forms of the outcome and exposure mechanisms
- the power to empirically identify more complex functional forms
  - Interactions
  - nonlinear, and higher-order relationships
  - broader range of functional forms (e.g., sinusoidal)

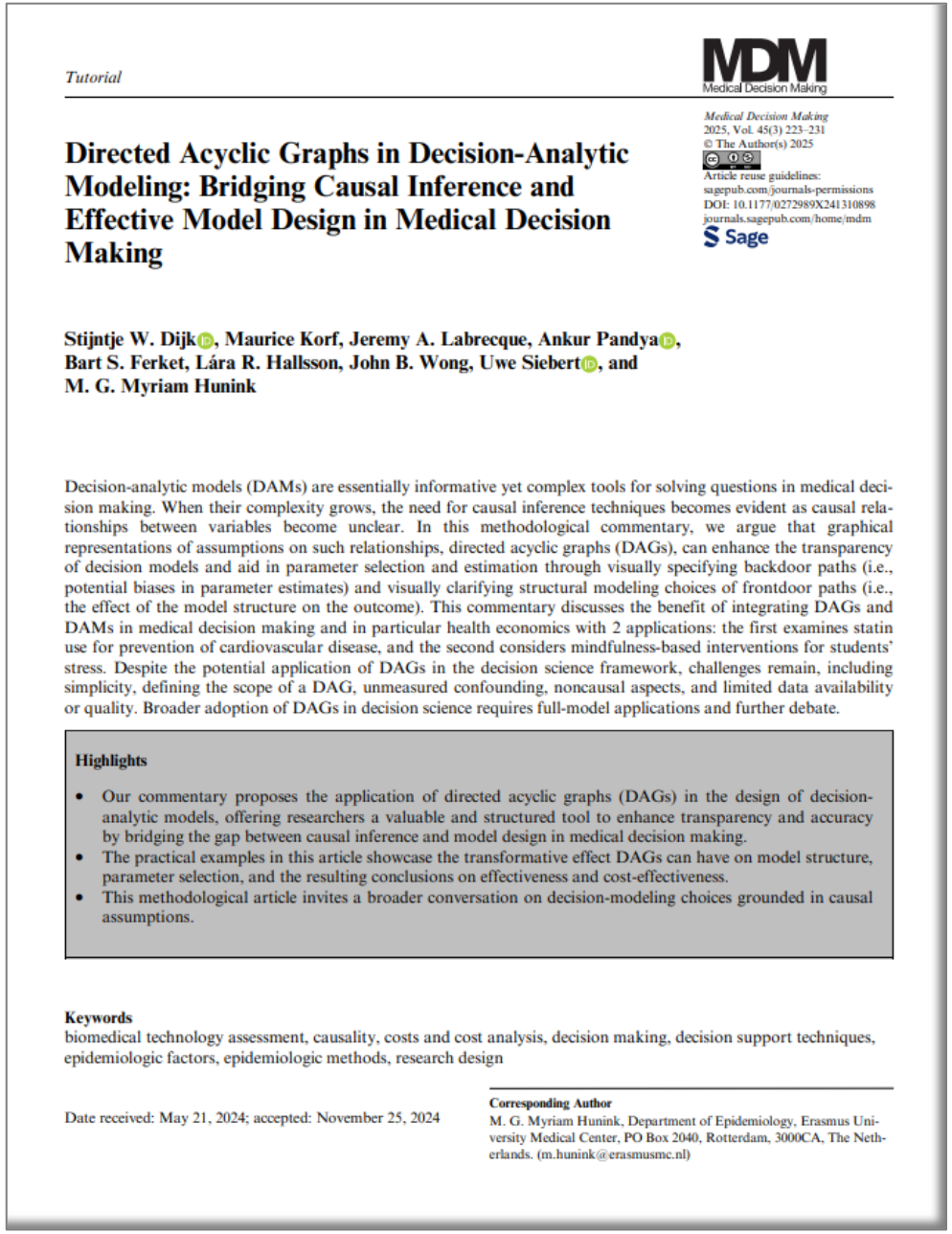
Support Clinical Guidelines and HTA:

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 …



Tutorial

MDM Medical Decision Making

Medical Decision Making
2025, Vol. 45(3) 223–231

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)

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.

UMIT TIROL

Causal Modeling

- Match your model structure with a causal diagram
- Check model input parameters for causality
- Do not base variable selection for regression equations on statistical significance but on DAGs
- Have appropriate methods for time-varying confounding been used?

*ISPOR-SMDM Modeling Good Research Practices Task Force

https://www.ispor.org/heor-resources/good-practices/report-other-report-types/-in-category/index-types/ispor-good-practices-for-outcomes-research

*Best Practice Causal Inference
ISPOR Good Research Practices for Retrospective Databases Analysis Task Force -- Part I-III


Volume 12 • Number 8 • 2009
VALUE IN HEALTH

Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III

Michael L. Johnson, PhD,[1] William Crown, PhD,[2] Bradley C. Martin, PharmD, PhD,[3] Colin R. Dormuth, MA, ScD,[4] Uwe Siebert, MD, MPH, MSc, ScD[5]

[1]University of Houston, College of Pharmacy, Department of Clinical Sciences and Administration, Houston, TX, USA; Houston Center for Quality of Care and Utilization Studies, Department of Veteran Affairs, Michael E. DeBakey VA Medical Center, Houston, TX, USA; [2]i3 Innovus, Waltham, MA, USA; [3]Division of Pharmaceutical Evaluation and Policy, College of Pharmacy, University of Arkansas for Medical Sciences, Little Rock, AR, USA; [4]Department of Anesthesiology, Pharmacology & Therapeutics, University of British Columbia; Pharmacoepidemiology Group, Therapeutics Initiative, Vancouver, BC, Canada; [5]Department of Public Health, Information Systems and Health Technology Assessment, UMIT, University of Health Sciences, Medical Informatics and Technology, Hall i.T., Austria; Institute for Technology Assessment and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, and Department of Health Policy and Management, Harvard School of Public Health, Boston, MA, USA

ABSTRACT

Objectives: Most contemporary epidemiologic studies require complex analytical methods to adjust for bias and confounding. New methods are constantly being developed, and older more established methods are yet appropriate. Careful application of statistical analysis techniques can improve causal inference of comparative treatment effects from nonrandomized studies using secondary databases. A Task Force was formed to offer a review of the more recent developments in statistical control of confounding.

Methods: The Task Force was commissioned and a chair was selected by the ISPOR Board of Directors in October 2007. This Report, the third in this issue of the journal, addressed methods to improve causal inference of treatment effects for nonrandomized studies.

Results: The Task Force Report recommends general analytic techniques and specific best practices where consensus is reached including: use of stratification analysis before multivariable modeling, multivariable regression including model performance and diagnostic testing, propensity scoring, instrumental variable, and structural modeling techniques including marginal structural models, where appropriate for secondary data. Sensitivity analyses and discussion of extent of residual confounding are discussed.

Conclusions: Valid findings of causal therapeutic benefits can be produced from nonrandomized studies using an array of state-of-the-art analytic techniques. Improving the quality and uniformity of these studies will improve the value to patients, physicians, and policymakers worldwide.

Keywords: causal inference, comparative effectiveness, nonrandomized studies, research methods, secondary databases.

Background to the Task Force

In September 2007, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Health Science Policy Council recommended that the issue of establishing a Task Force to recommend Good Research Practices for Designing and Analyzing Retrospective Databases be considered by the ISPOR Board of Directors. The Council's recommendations concerning this new Task Force were to keep an overarching view toward the need to ensure internal validity and improve causal inference from observational studies, review prior work from past and ongoing ISPOR task forces and other initiatives to establish baseline standards from which to set an agenda for work. The ISPOR Board of Directors approved the creation of the Task Force in October 2007. Task Force leadership and reviewer groups were finalized by December 2007, and the first teleconference took place in January 2008.

Address correspondence to: Michael L. Johnson, University of Houston, College of Pharmacy, Department of Clinical Sciences and Administration, Houston, TX 77030, USA. E-mail: mikejohnson@uh.edu

10.1111/j.1524-4733.2009.00602.x

Task Force members were experienced in medicine, epidemiology, biostatistics, public health, health economics, and pharmacy sciences, and were drawn from industry, academia, and as advisors to governments. The members came from the UK, Germany, Austria, Canada, and the United States.

Beginning in January 2008, the Task Force conducted monthly teleconferences to develop core assumptions and an outline before preparing a draft report. A face-to-face meeting took place in October 2008, to develop the draft, and three forums took place at the ISPOR Meetings to develop consensus for the final draft reports. The draft reports were posted on the ISPOR website in May 2009 and the task forces' reviewer group and ISPOR general membership were invited to submit their comments for a 2 week reviewer period. In total, 38 responses were received. All comments received were posted to the ISPOR website and presented for discussion at the Task Force forum during the ISPOR 12th Annual International Meeting in May 2009. Comments and feedback from the forum and reviewer and membership responses were considered and acknowledged in the final reports. Once consensus was reached, the manuscript was submitted to *Value in Health*.

 1098-3015/09/** **-**

1. Berger et al. … defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources … Part I., Value in Health 2009
2. Cox et al., … approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources … Part II, Value in Health 2009
3. Johnson et al., … analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources … Part III, Value in Health 2009

*Best Practice Causal Inference
ISPOR Good Research Practices for Retrospective Databases Analysis Task Force -- Part I-III

Recommendations Part III (Johnson et al, ViH 2009):

- "Include variables that are only weakly related to treatment selection because they may potentially reduce bias more than they increase variance."
- "All factors that are theoretically related to outcome or treatment selection should be included despite statistical significance at traditional levels of significance."
- "In the presence of time-varying confounding, standard statistical methods may be biased, and alternative methods such as marginal structural models or g-estimation should be examined."

*Causal Inference and Decision-Analytic Modeling

- "Causal inference methods aim for drawing causal conclusions from empirical data on the relationship of pre-specified interventions on a specific target outcome and apply a counterfactual framework and statistical techniques to derive causal effects of exposures or interventions from these data."
- "Decision analyses are based on decision-analytic models to mimic the course of disease as well as aspects and consequences of the intervention in order to quantitatively optimize the decision."

Kühne F, Schomaker M, Stojkov I, Jahn B, Conrads-Frank A, Siebert S, Sroczynski G, Puntscher S, Schmid D, Schnell-Inderst P, Siebert U. Causal evidence in health decision making: methodological approaches of causal inference and health decision science. Ger Med Sci. 2022;20:Doc12. doi: 10.3205/000314.

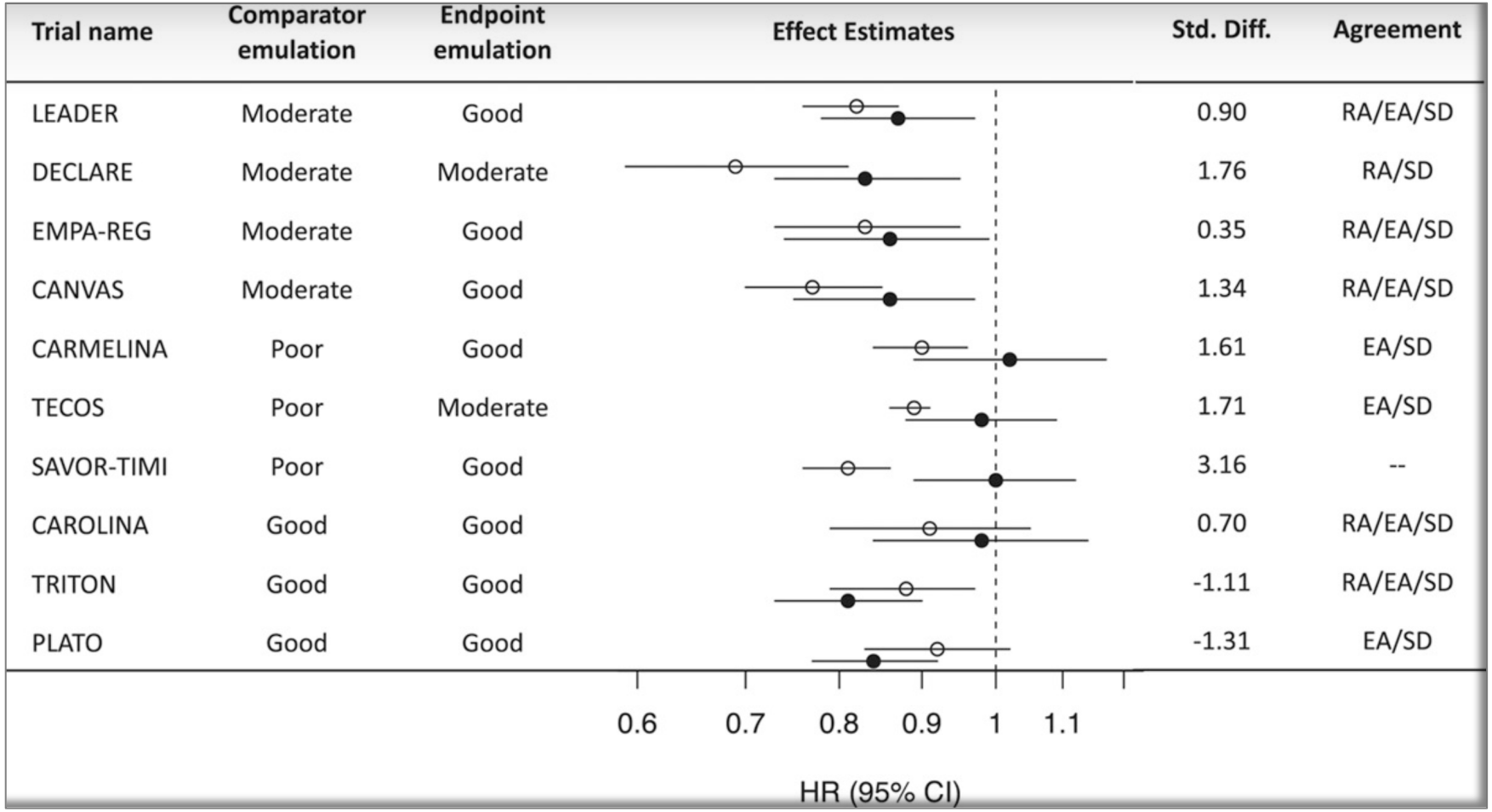
*Microsimulation: UKPDS Outcomes Model

Table 2. Sample size, functional form, parameters and beta coefficients (SEs) for seven equations to estimate the probability of diabetes-related complications

	Eq. 1	Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7
Complication	IHD	MI	CHF	STROKE	AMP	BLIND	RENAL
No. of subjects	3612	3642	3607	3607	3642	3642	3642
Functional form	Weibull	Weibull	Weibull	Weibull	Weibull	Weibull	Weibull
Parameters	Estimate of coefficient (SE)						
λ	−5.310 (0.174)	−4.977 (0.160)	−8.018 (0.408)	−7.163 (0.342)	−8.718 (0.613)	−6.464 (0.326)	−10.016 (0.939)
ρ	1.150 (0.067)	1.257 (0.060)	1.711 (0.158)	1.497 (0.126)	1.451 (0.232)	1.154 (0.121)	1.865 (0.387)
AGE	0.031 (0.008)	0.055 (0.006)	0.093 (0.016)	0.085 (0.014)		0.069 (0.014)	
FEMALE	−0.471 (0.143)	−0.826 (0.103)		−0.516 (0.171)			
AC		−1.312 (0.341)					
SMOK		0.346 (0.097)		0.355 (0.179)			
BMI			0.066 (0.017)				
HBA1C	0.125 (0.035)	0.118 (0.025)	0.157 (0.057)	0.128 (0.042)	0.435 (0.066)	0.221 (0.050)	
SBP	0.098 (0.037)	0.101 (0.026)	0.114 (0.056)	0.276 (0.042)	0.228 (0.075)		0.404 (0.106)
TOTAL:HDL				0.113 (0.025)			
Ln (TOTAL:HDL)	1.498 (0.202)	1.190 (0.169)					
PVD					2.436 (0.521)		
ATRFIB				1.428 (0.472)			
IHD		0.914 (0.150)					
CHF		1.558 (0.202)		1.742 (0.287)			
BLIND					1.812 (0.462)		2.082 (0.551)

Biases and Benchmarking in Real-World Case Examples

RCT DUPLICATE – Benchmarking RWE against RCTs

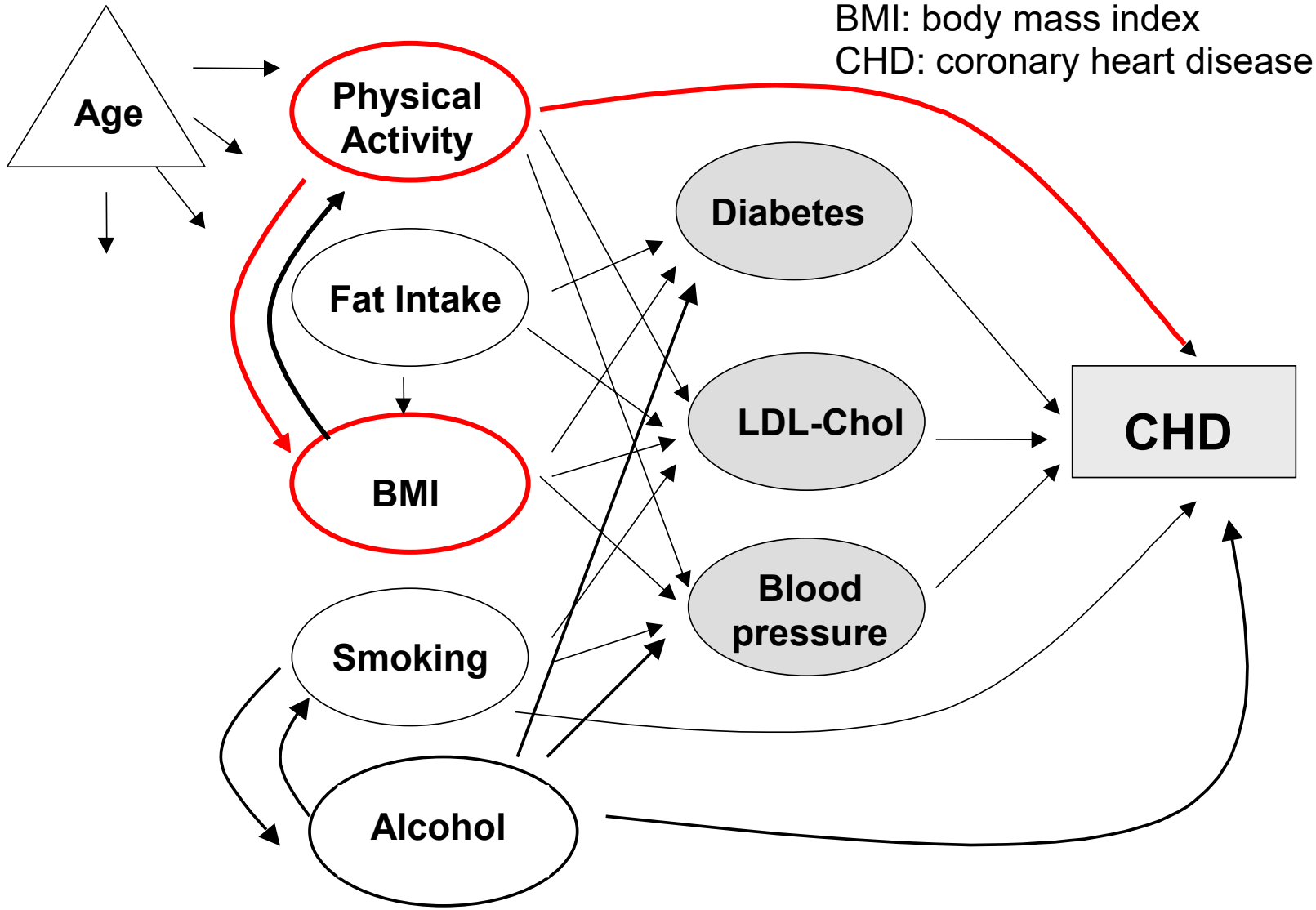


Franklin JM et al., Circulation. 2021 doi: 10.1161/CIRCULATIONAHA.120.051718

Case Examples

- Parametric g-Formula
  - First application of parametric g-formula: assessment of cardiologic risk factors (see appendix)
- Marginal structural model (MSM) with inverse-probability-of-treatment weighting (IPTW)
  - Pivotal re-analysis of HIV treatment with MACS registry data (see appendix)
- Clone-Censor-Weight (CCW) Approach with inverse-probability-of-censoring weighting (IPCW)
  - 2nd line treatment in oncology (see main slides)

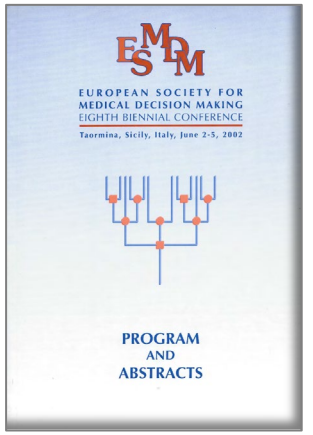
Case Example g-Formula: WHO Project Cardiology



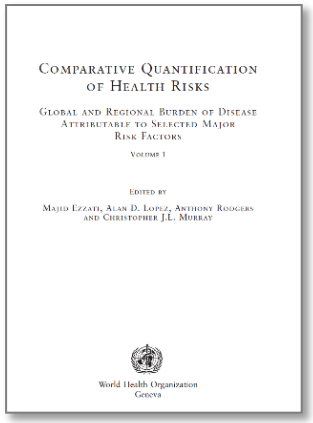
Parametric g-Formula

Case Example: WHO-Project "Causal CHD Web"

Siebert U, Hernán MA, Robins JM. Monte Carlo simulation of the direct and indirect impact of risk factor interventions on coronary heart disease. An application of the g-formula. Proceedings of the 8th Biennial Conference of the European Society for Medical Decision Making in Taormina, Sicily, Italy. June 2-5, 2002:p51.



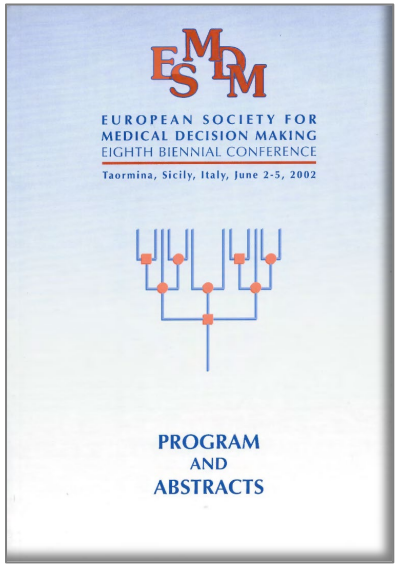
Robins JM, Hérnan MA, Siebert U. Estimations of the effects of multiple interventions. In: Ezzati M, Lopes AD, Rodgers A, Murray CJL, eds. Comparative quantification of health risks; global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization, 2004:2191-2230.



This example represents the first application of the parametric g-formula in the literature

Abstract

First Application of Parametric g-Formula



Siebert, Robins, Hernán, ESMDM 2002

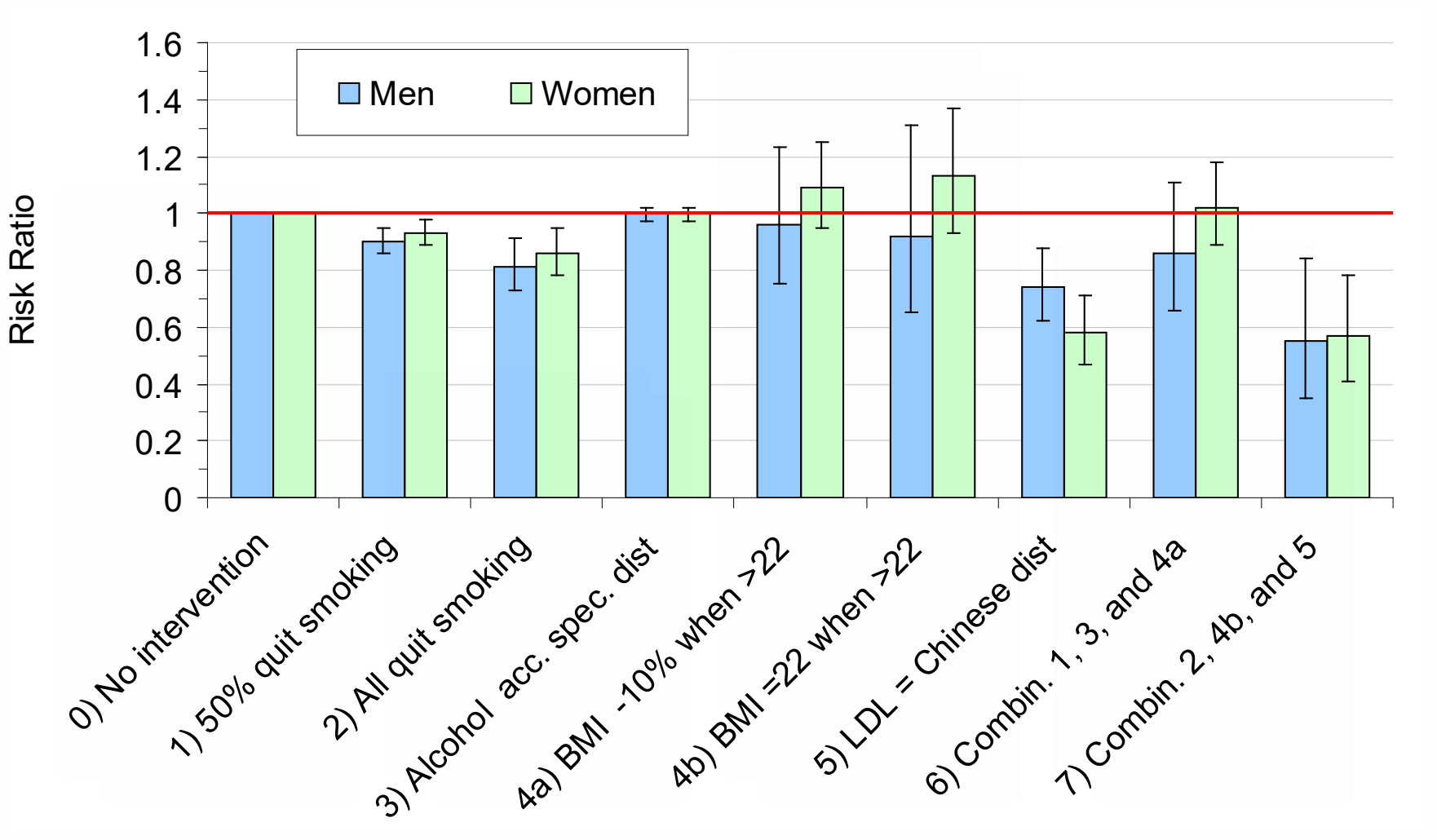
https://cdn1.sph.harvard.edu/wp-content/uploads/sites/604/2022/04/Siebert_2002_ProceedingsESMDM_Impact_interventions_CHD_application_g-formula.pdf

MONTE CARLO SIMULATION OF THE DIRECT AND INDIRECT IMPACT OF RISK FACTOR INTERVENTIONS ON CORONARY HEART DISEASE. AN APPLICATION OF THE G-FORMULA

Uwe Siebert (1), Miguel A. Hernán (2), James M. Robins (2)
(1) Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, MA, USA;
(2) Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Background: WHO has established a project on comparative risk assessment for coronary heart disease (CHD) which addresses the overall impact of several public health interventions on the risk of CHD. As there are multiple direct and indirect risk factors that are part of the causal web of CHD, such an evaluation has to consider not only the direct effect of the risk factors under intervention but has to include the effect mediated through other risk factors. As various risk factors simultaneously act as confounders and as intermediate steps, traditional regression analysis fails and alternative methods have to be used. Our goal was to evaluate the impact of several interventions on CHD risk factors using the g-formula.
Methods: We used data of the Framingham Offspring Study, in which risk factor and disease status of 5124 subjects was assessed in one baseline and 4 follow-up exams over 20 years. Based on a causal diagram which represented prior knowledge about the causal links among CHD risk factors, potential risk factors, confounders, and intermediate variables were defined. We used the g-formula to estimate the counterfactual CHD risk in the Framingham Offspring data under each intervention. The g-formula (Robins, 1986) is a general nonparametric method that allows to compute counterfactual proportions under the assumption of no unmeasured confounders. In a first step, we fit pooled logistic regression models to predict risk factor distributions and the risk of CHD based on the given risk factor history. In a second step, Monte Carlo technique was used to simulate a sample of 10000 counterfactual subjects and their experience regarding risk factors and CHD. Bootstrap methods were used to estimate 95% confidence intervals (CI) of the counterfactual CHD risks and relative risks (RR). All analyses were performed separately by gender. Evaluated strategies included interventions on smoking, alcohol consumption, body mass index (BMI), low density lipoprotein (LDL), and a combined strategy.
Results: After exclusions, our analyses included 2230 men (47.8%) and 2440 (52.2%) women with 189 and 68 CHD events, respectively. The observed 12-year risk of CHD in the study population was 8.47% (CI: 7.37%-9.73%) for males and 2.79% (CI: 2.19%-3.54%) for females. The simulated 12-year risk of CHD under no intervention was 8.48% for males (CI: 6.72%-10.24%) and 2.82% for females (CI: 2.10%-3.54%). Smoking cessation at baseline in all male and female smokers had a relative risk of 0.80 (CI: 0.70-0.91) and 0.83 (CI: 0.70-1.00), respectively. The relative risk after shifting the LDL distribution to the distribution of the Chinese population was 0.68 (CI: 0.52-0.89) for men and 0.48 (CI: 0.35-0.67) for women. Shifting alcohol consumption to moderate alcohol intake or constantly lowering BMI to 22kg/m^2 did not change CHD risk significantly. The combined intervention on smoking cessation, BMI, and LDL reduced the CHD risk by 53% (RR: 0.47; CI: 0.32-0.69) in men and by 61% (RR: 0.39; CI: 0.23-0.67) in women.
Conclusions: The g-formula could be applied in a situation where traditionl regression analysis could not be used, because risk factor variables were confounders as well as intermediate steps. Highest risk reductions could be achieved by interventions on smoking cessation and LDL. Combined interventions may reduce CHR risk by more than 50%.

Results


Risk Ratio
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Men
Women
0) No intervention
1) 50% quit smoking
2) All quit smoking
3) Alcohol acc. spec. dist
4a) BMI -10% when >22
4b) BMI =22 when >22
5) LDL = Chinese dist
6) Combin. 1, 3, and 4a
7) Combin. 2, 4b, and 5

Case Example MSM with IPW: HIV Treatment

Hernan MA. Brumback B. Robins JM.

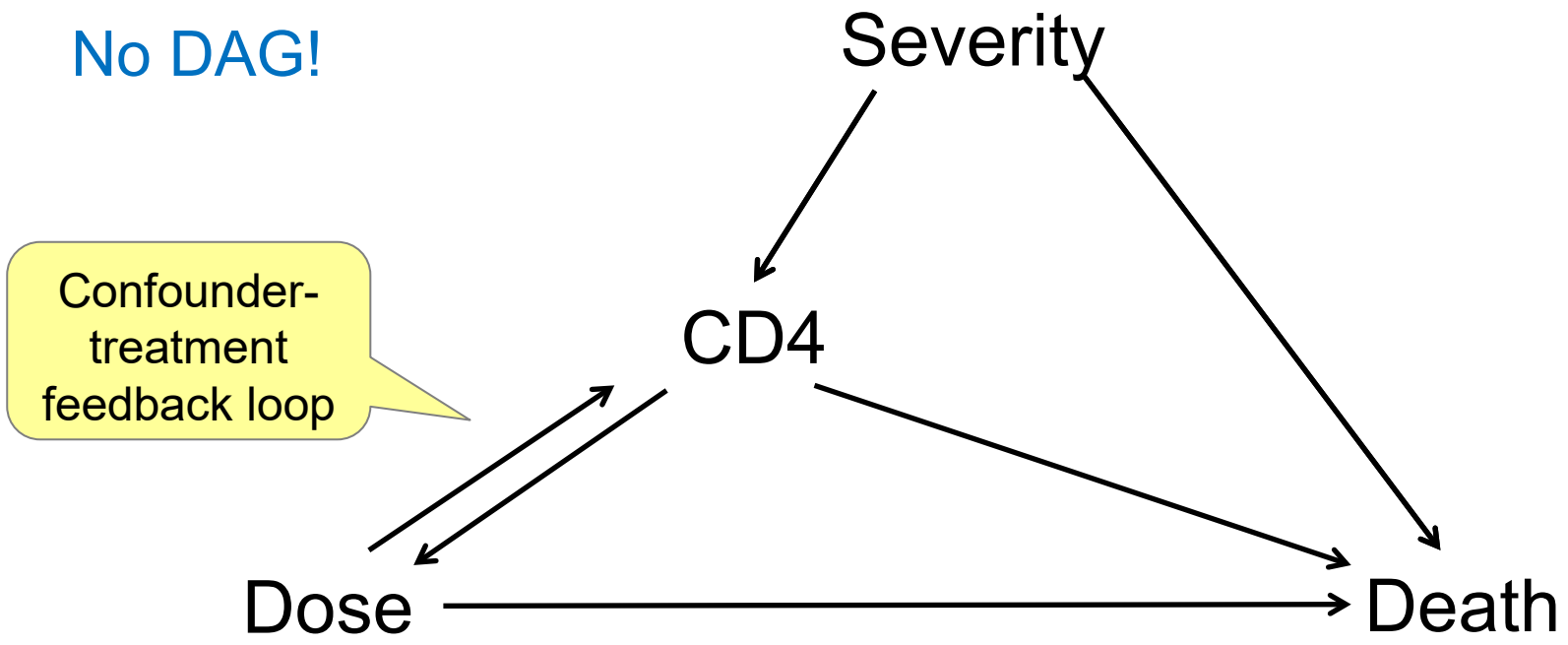
Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men.

Epidemiology. 11(5):561-70, 2000

- Data from Multicenter AIDS Cohort Study (MACS)
- Causal effect of antiviral drug (dose) on mortality

Case Example MSM with IPTW: HIV Treatment

No DAG!



Case Example MSM with IPTW: HIV Treatment

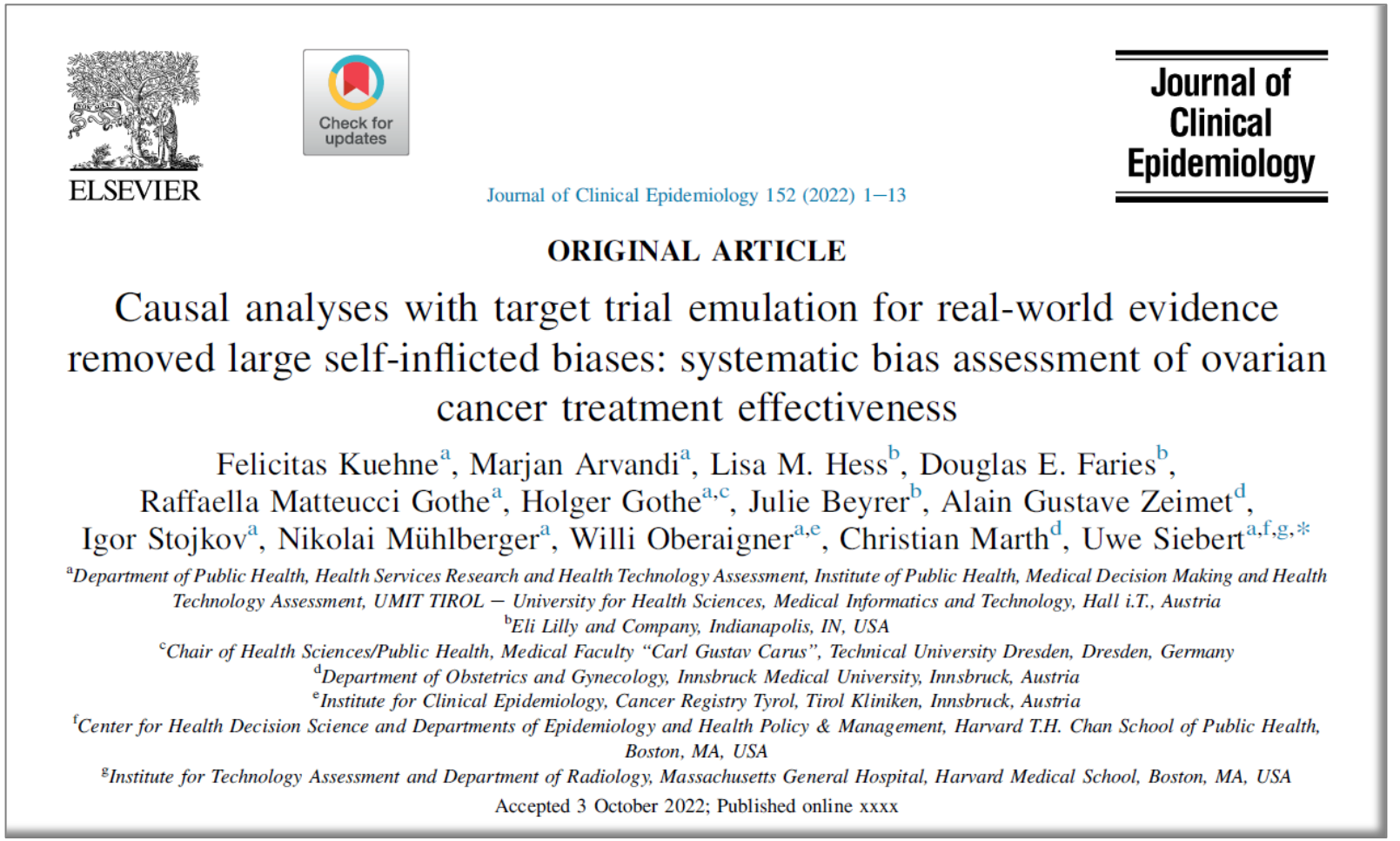
Analysis (high vs. low dose drug)	RR	Sign.
Not adjusted for CD4	3.6	p<0.05
Adjusted for CD4	2.3	p<0.05
Marginal structural model (adjusting for time-dependent CD4)	0.7	p<0.05

Data from Hernan MA, Brumback B, Robins JM. Epidemiology 2000

Case Examples

- Parametric g-Formula
  - First application of parametric g-formula: assessment of cardiologic risk factors (see appendix)
- Marginal structural model (MSM) with inverse-probability-of-treatment weighting (IPTW)
  - Pivotal re-analysis of HIV treatment with MACS registry data (see appendix)
- TTE using the clone-censor-weight (CCW) approach with inverse-probability-of-censoring weighting (IPCW)
  - 2nd line treatment in oncology
  - Should we benchmark RWE against RCTs?

Case Example Clone-Censor-Weight: Oncology



Journal of Clinical Epidemiology 152 (2022) 1–13

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Causal analyses with target trial emulation for real-world evidence removed large self-inflicted biases: systematic bias assessment of ovarian cancer treatment effectiveness

Felicitas Kuehne[a], Marjan Arvandi[a], Lisa M. Hess[b], Douglas E. Faries[b], Raffaella Matteucci Gothe[a], Holger Gothe[a,c], Julie Beyrer[b], Alain Gustave Zeimet[d], Igor Stojkov[a], Nikolai Mühlberger[a], Willi Oberaigner[a,e], Christian Marth[d], Uwe Siebert[a,f,g,*]

[a]*Department of Public Health, Health Services Research and Health Technology Assessment, Institute of Public Health, Medical Decision Making and Health Technology Assessment, UMIT TIROL – University for Health Sciences, Medical Informatics and Technology, Hall i.T., Austria*

[b]*Eli Lilly and Company, Indianapolis, IN, USA*

[c]*Chair of Health Sciences/Public Health, Medical Faculty "Carl Gustav Carus", Technical University Dresden, Dresden, Germany*

[d]*Department of Obstetrics and Gynecology, Innsbruck Medical University, Innsbruck, Austria*

[e]*Institute for Clinical Epidemiology, Cancer Registry Tyrol, Tirol Kliniken, Innsbruck, Austria*

[f]*Center for Health Decision Science and Departments of Epidemiology and Health Policy & Management, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[g]*Institute for Technology Assessment and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA*

Accepted 3 October 2022; Published online xxxx

Kuehne et al., J Clinical Epidemiol, 2022

Reference Case RCT

Early versus delayed treatment of relapsed ovarian cancer (MRC OV05/EORTC 55955): a randomised trial

*Gordon J S Rustin, Maria E L van der Burg, Clare L Griffin, David Guthrie, Alan Lamont, Gordon C Jayson, Gunnar Kristensen, César Mediola, Corneel Coens, Wendi Qian, Mahesh K B Parmar, Ann Marie Swart, for the MRC OV05 and EORTC 55955 investigators**

Summary

Background Serum CA125 concentration often rises several months before clinical or symptomatic relapse in women with ovarian cancer. In the MRC OV05/EORTC 55955 collaborative trial, we aimed to establish the benefits of early treatment on the basis of increased CA125 concentrations compared with delayed treatment on the basis of clinical recurrence.

Methods Women with ovarian cancer in complete remission after first-line platinum-based chemotherapy and a normal CA125 concentration were registered for this randomised controlled trial. Clinical examination and CA125 measurement were done every 3 months. Patients and investigators were masked to CA125 results, which were monitored by coordinating centres. If CA125 concentration exceeded twice the upper limit of normal, patients were randomly assigned (1:1) by minimisation to early or delayed chemotherapy. Patients and clinical sites were informed of allocation to early treatment, and treatment was started as soon as possible within 28 days of the increased CA125 measurement. Patients assigned to delayed treatment continued masked CA125 measurements, with treatment commencing at clinical or symptomatic relapse. All patients were treated according to standard local practice. The primary outcome was overall survival. Analysis was by intention to treat. This study is registered, ISRCTN87786644.

Findings 1442 patients were registered for the trial, of whom 529 were randomly assigned to treatment groups and were included in our analysis (265 early, 264 delayed). With a median follow-up of 56·9 months (IQR 37·4–81·8) from randomisation and 370 deaths (186 early, 184 delayed), there was no evidence of a difference in overall survival between early and delayed treatment (HR 0·98, 95% CI 0·80–1·20, p=0·85). Median survival from randomisation was 25·7 months (95% CI 23·0–27·9) for patients on early treatment and 27·1 months (22·8–30·9) for those on delayed treatment.

Interpretation Our findings showed no evidence of a survival benefit with early treatment of relapse on the basis of a raised CA125 concentration alone, and therefore the value of routine measurement of CA125 in the follow-up of patients with ovarian cancer who attain a complete response after first-line treatment is not proven.

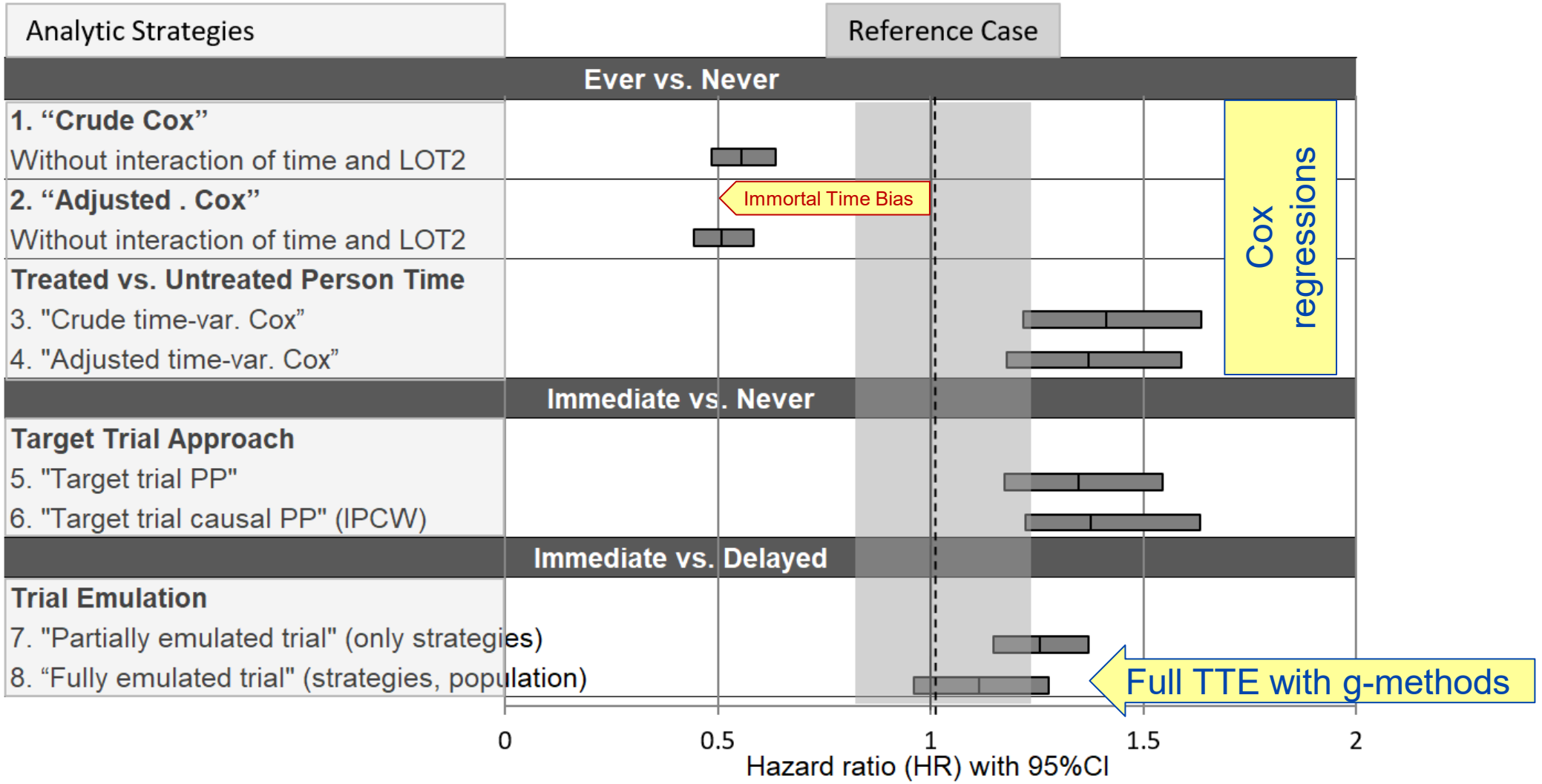
Rustin et al.; Lancet 2010

	Hazard ratio (95% CI)	Log-rank p value
Unadjusted	0·98 (0·80–1·20)	0·85
Adjusted		
For stratification factors*	0·99 (0·80–1·22)	--
For prognostic factors†	0·98 (0·79–1·21)	--
For stratification and prognostic factors	1·01 (0·82–1·25)	--
Sensitivity analyses‡	1·01 (0·82–1·23)	0·96

*Age, International Federation of Gynecology and Obstetrics stage, first-line chemotherapy, time from completion of first-line chemotherapy to doubling of CA125 concentration, and country. †Histology, WHO performance status, and time from doubling of CA125 concentration to randomisation. ‡Sensitivity analyses of non-curtailed data (all follow-up data received, not curtailed at 5 years for MRC OV05 and 3 years for EORTC 55955).

Table 3: **Hazard ratios for overall survival**

Bias Assessment

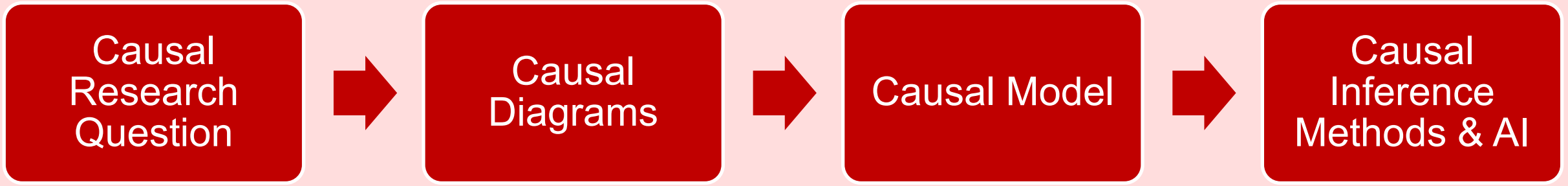


Kuehne F et al., J Clin Epidemiol, 2022

Summary

- Combining visual, structural and statistical techniques helps avoiding biases in causal inference from observational studies
- Causal graphs are helpful tools in causal inference
- Traditional methods (multivariate regression, propensity score etc.) can control for time-independent (baseline) confounding
- g-methods/TMLE must be used to control for time-varying confounding
- The target trial approach should be applied to any observational study to avoid self-inflicted biases
- Machine learning can be helpful to specify causal functional forms
- Causal approaches must be integrated in decision-analytic modeling
- Benchmarking is important to gain trust in RWE
- HTA increases the use of RWE, target trial emulation and causal inference
- AI plays an increasing role in causal modeling
- Join us in disseminating causal inference tools in research, education and policy advice

*Role of AI in Causal Modeling



Meaningful and well-defined interventions

Causal natural history and intervention effects

Causal structure of events, transitions etc.

Regression, Machine learning/AI, g-Methods, Target Trial Emulation

Vision - Role of generative AI:
Asking questions that we cannot ask, as we don't understand (yet).

Role of AI:
Causal discovery = algorithms to learn the DAG from the data (e.g., PC algorithm)

Role of AI:
Use of LLM + user interface to generate decision models for HEOR

Role of AI:
Machine learning in data analysis to learn the form of the causal influence function, identification of effect modification

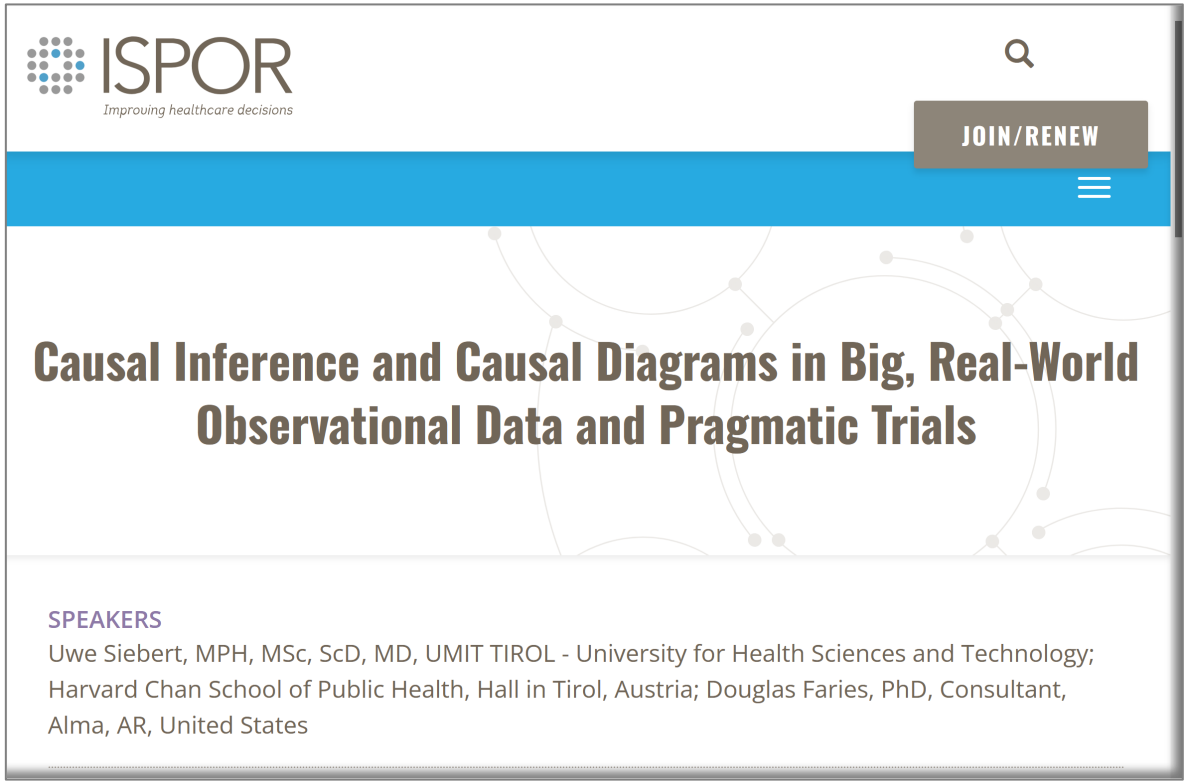
Outlook

Further Causal Topics of Interest for NANA

- DAG development
  - Practical aspects in Delphi panels
- Dynamic target trial protocol
  - Adjustment of the TT protocol as available data are revealed
- Transportability
  - Recent overview: Adamson B et al. Pharmacoepidemiol Drug Saf. 2026 June, https://onlinelibrary.wiley.com/doi/10.1002/pds.70396, HTAi 2026 panel: “Helping or Hurting Health Technology Assessments? When is Transportability An Appropriate Solution When Local Real-World Evidence is Limited?”
- Causal Machine Learning and Causal AI
  - Functional forms vs. variable selection, AI (still) only under human control
- Indirect treatment comparisons (ITC) and meta-analysis
  - Use of DAGs and TTE in ITC, e.g. population-adjusted indirect comparisons (PAIC) with individual data , see presentation at HTAi
- The data!
  - Quality, completeness, access, linkage

Want to Hear More about Causal Inference Medical Decision Making?

https://smdm.org/causal-diagrams-target-trial-emulation-and-causal-inference-for-modeling-and-medical-decision-making/



https://www.ispor.org/heor-resources/presentations-database/session-cti/ispor-2026/causal-inference-and-causal-diagrams-in-big-real-world-observational-data-and-pragmatic-trials-2/

Continuing Education www.htads.org

This presentation contains parts of the more comprehensive UMIT TIROL Continuing Education Program www.htads.org

Including a 5-day course on Causal Inference,
10 - 14 May 2027

www.htads.org

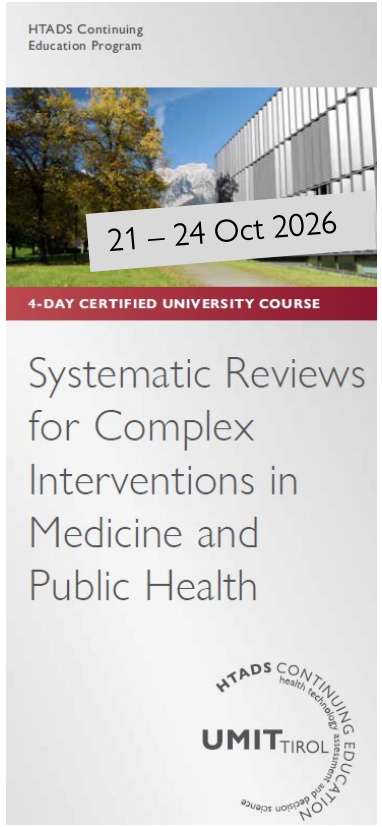
Video: https://www.youtube.com/watch?v=QHEXTI_L9dw

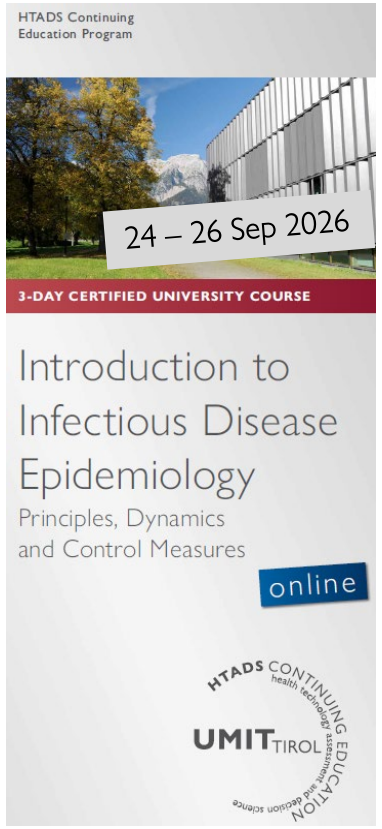
Scan!

HTADS Continuing Education Program – University Certified Courses

HTADS Continuing Education Program Certified Courses

Information: Web: www.htads.org, Email: htads@umit-tirol.at

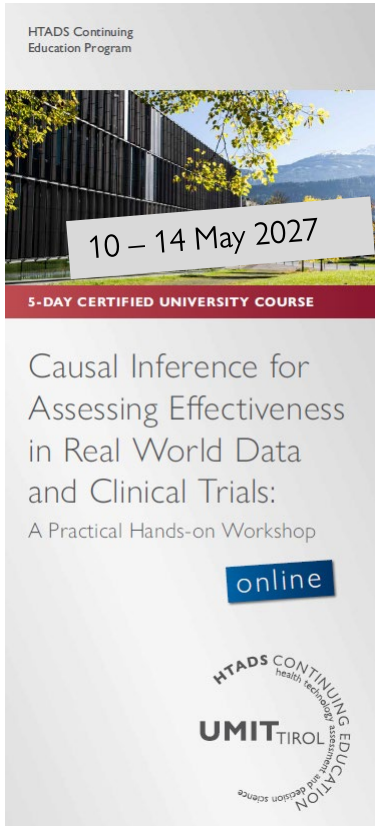


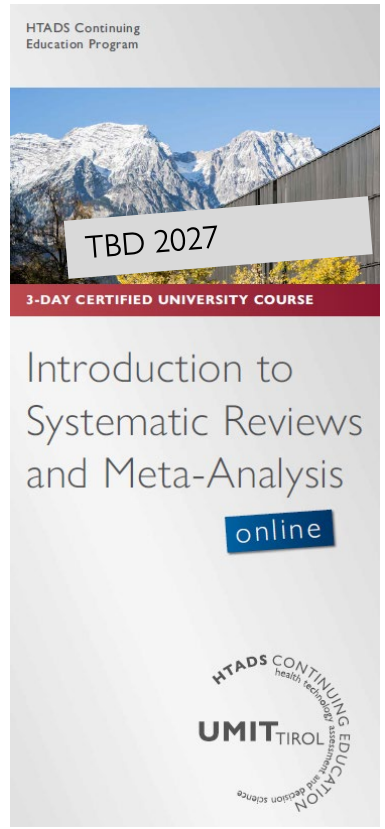


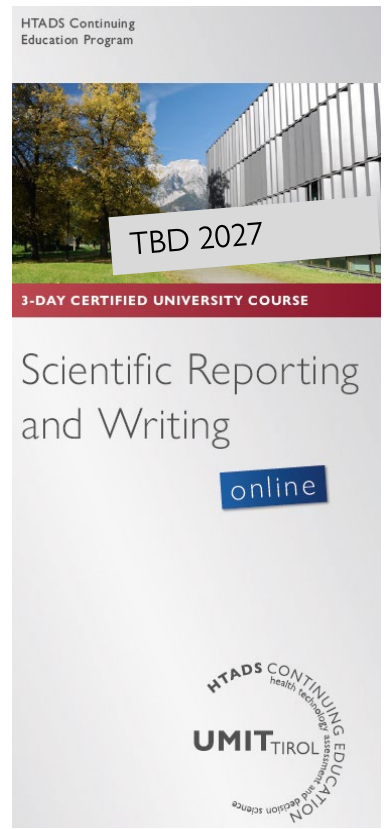

HTADS Continuing Education Program
25 – 27 Jan 2027
3-DAY CERTIFIED UNIVERSITY COURSE
Introduction to Health Technology Assessment and Health Economics
online
UMIT TIROL

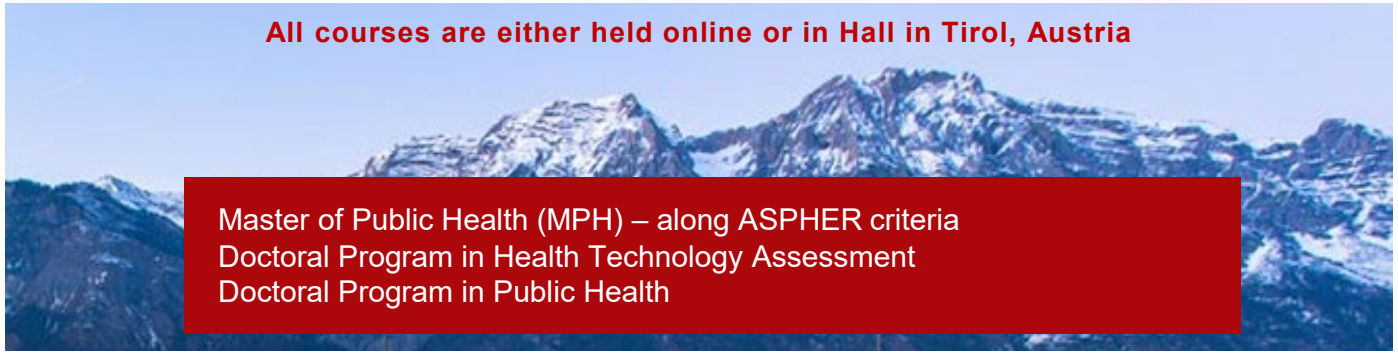

HTADS Continuing Education Program
15 - 19 Feb 2027
5-DAY CERTIFIED UNIVERSITY COURSE
Winter School in Clinical Epidemiology
UMIT TIROL


HTADS Continuing Education Program
21 – 23 Apr 2027
3-DAY CERTIFIED UNIVERSITY COURSE
Modeling Approaches for HTA
A Practical Hands-on Workshop
Six Model Types – Four Software Options – Ready to Take the Challenge?
UMIT TIROL


HTADS Continuing Education Program
10 – 14 May 2027
5-DAY CERTIFIED UNIVERSITY COURSE
Causal Inference for Assessing Effectiveness in Real World Data and Clinical Trials:
A Practical Hands-on Workshop
online
UMIT TIROL


HTADS Continuing Education Program
TBD 2027
3-DAY CERTIFIED UNIVERSITY COURSE
Introduction to Systematic Reviews and Meta-Analysis
online
UMIT TIROL


HTADS Continuing Education Program
TBD 2027
3-DAY CERTIFIED UNIVERSITY COURSE
Scientific Reporting and Writing
online
UMIT TIROL

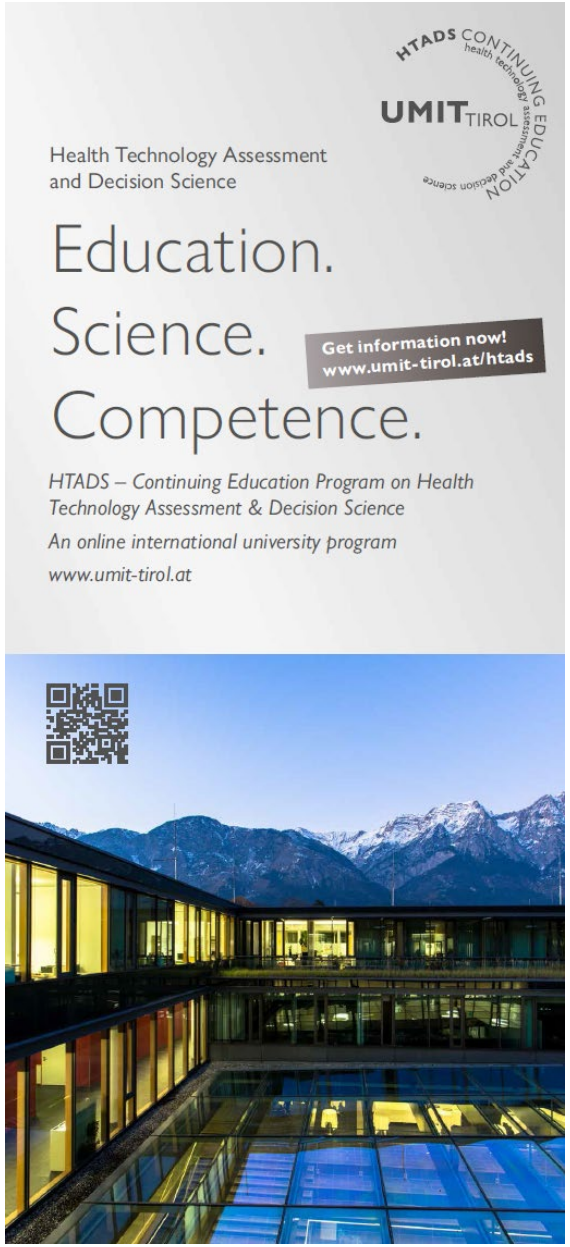

All courses are either held online or in Hall in Tirol, Austria
Master of Public Health (MPH) – along ASPHER criteria
Doctoral Program in Health Technology Assessment
Doctoral Program in Public Health

HTADS – Info Letter

If you would like to receive additional material, please send an email to:

htads@umit-tirol.at

Subject: Subscribe HTADS



Thank you!

Questions? Contact me:

@uweSiebert9

uwe-siebert9

uwe.siebert@umit-tirol.at

Web: www.htads.org