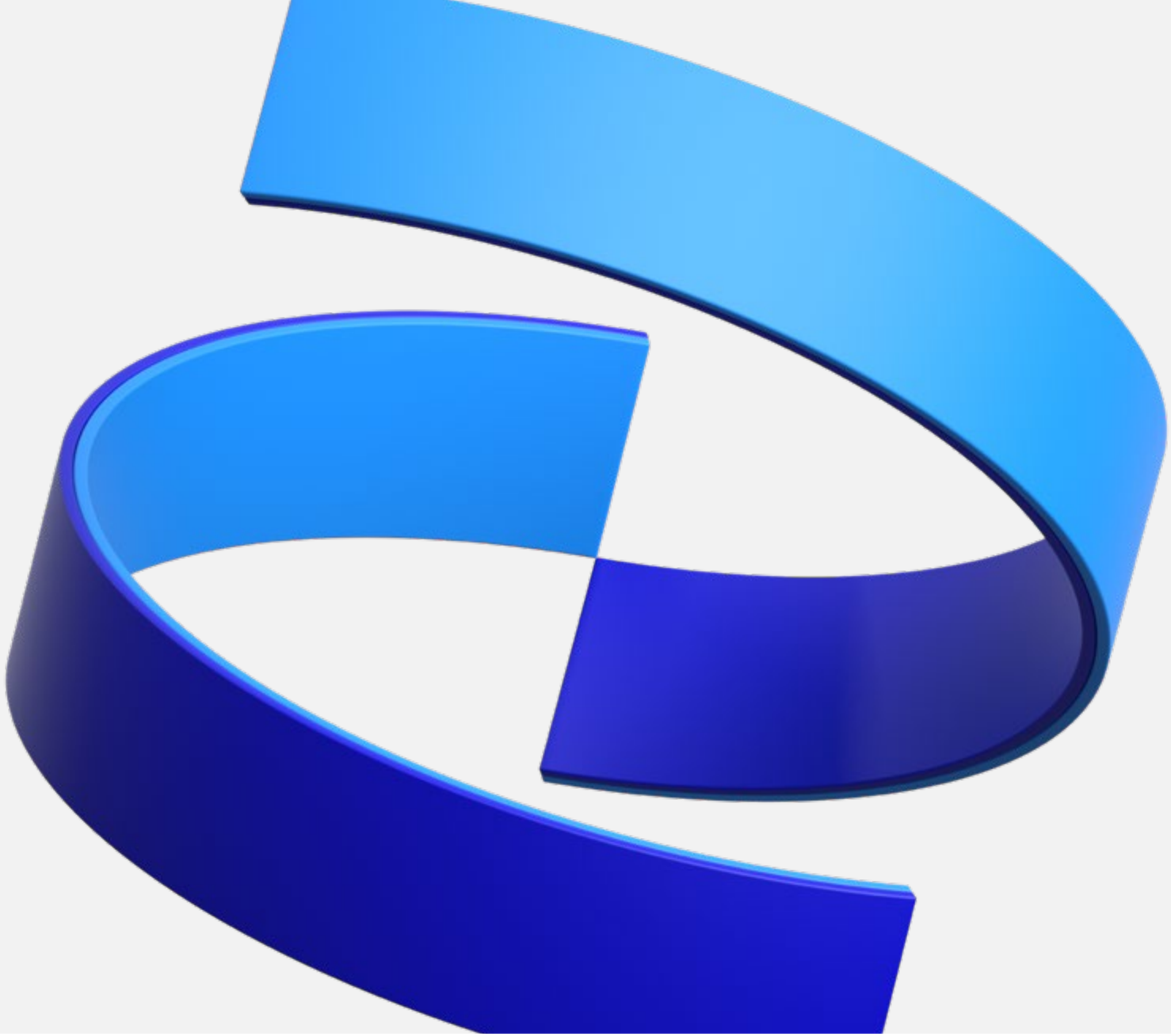
Causal Roadmap for Regulatory Real-World Evidence Studies



Martin Ho, FASA, MS

December 6, 2023

Acknowledgement

This presentation is the result of a close collaboration of Team 3 members of the ASA Biopharmaceutical Section RWE Scientific Working Group

- David Benkeser (Emory)
- Yixin Fang (Abbvie) - team co-lead
- Douglas Faries (Eli Lilly)
- Susan Gruber (Putnam Data Sciences, LLC)
- Martin Ho (Pfizer) - team co-lead
- Mark van der Laan (University of California Berkeley)
- Pallavi Mishra-Kalyani (FDA Center for Drugs Evaluation and Research)

Disclaimer

The opinions stated here are my own, not those of my employer.

Publications

Research Article

Examples of Applying RWE Causal-Inference Roadmap to Clinical Studies

Martin Ho, Susan Gruber, Yixin Fang, Douglas E. Faris, Pallavi Mishra-Kalyani, David Benkeser &
...show all

Received 02 Jun 2022, Accepted 25 Jan 2023, Published online: 21 Mar 2023

Cite this article https://doi.org/10.1080/19466315.2023.2177333 Check for updates

Special Section: A Collection of Articles on Opportunities and Challenges in Utilizing Real-World Data for Clinical Trials and Medical Product Development

The Current Landscape in Biostatistics of Real-World Data and Evidence: Causal Inference Frameworks for Study Design and Analysis

Martin Ho, Mark van der Laan, Hana Lee, Jie Chen, Kwan Lee, Yixin Fang, ...show all

Pages 43-56 | Received 19 May 2020, Accepted 26 Jan 2021, Published online: 15 Mar 2021

Cite this article https://doi.org/10.1080/19466315.2021.1883475 Check for updates

The latest iteration of causal roadmap

Journal of Clinical and Translational Science

www.cambridge.org/cts

Research Methods and Technology Review Article

Cite this article: Dang LE, Gruber S, Lee H, Dahabreh IJ, Stuart EA, Williamson BD, Wyss R, Díaz I, Ghosh D, Kıcıman E, Alemayehu D, Hoffman KL, Vossen CY, Huml RA, Ravn H, Kvist K, Pratley R, Shih M-C, Pennello G, Martin D, Waddy SP, Barr CE, Akacha M, Buse JB, van der Laan M, and Petersen M. A causal roadmap for generating high-quality real-world evidence. *Journal of Clinical and Translational Science* **7**: e212, 1–12. doi: 10.1017/cts.2023.635

Received: 10 May 2023
Revised: 1 September 2023
Accepted: 17 September 2023

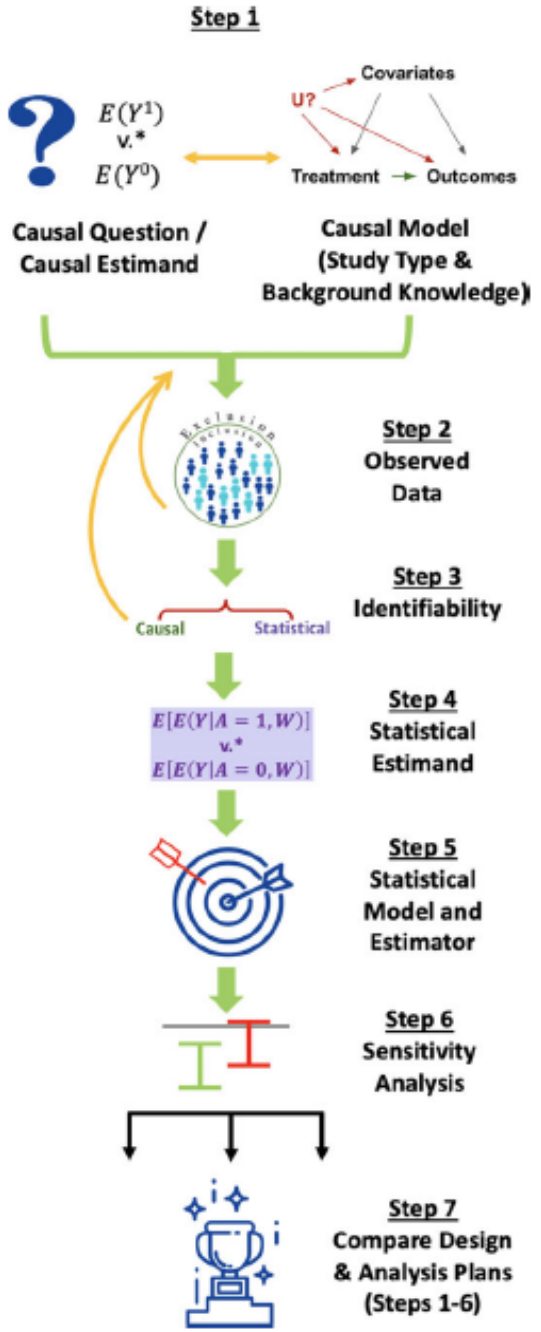
Keywords:
Causal inference; real-world evidence; sensitivity analysis; simulations; estimands; machine learning

Corresponding author:
L. E. Dang, MD, MPH;
Email: lauren.eyler@berkeley.edu

A causal roadmap for generating high-quality real-world evidence

Lauren E. Dang[1], Susan Gruber[2], Hana Lee[3], Issa J. Dahabreh[4], Elizabeth A. Stuart[5], Brian D. Williamson[6], Richard Wyss[7], Iván Díaz[8], Debashis Ghosh[9], Emre Kıcıman[10], Demissie Alemayehu[11], Katherine L. Hoffman[12], Carla Y. Vossen[13], Raymond A. Huml[14], Henrik Ravn[15], Kajsa Kvist[15], Richard Pratley[16], Mei-Chiung Shih[17,18], Gene Pennello[19], David Martin[20], Salina P. Waddy[21], Charles E. Barr[22,23], Mouna Akacha[24], John B. Buse[25], Mark van der Laan[1] and Maya Petersen[1]

[1]Department of Biostatistics, University of California, Berkeley, CA, USA; [2]TL Revolution, Cambridge, MA, USA; [3]Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA; [4]CAUSALab, Department of Epidemiology and Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [5]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [6]Biostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA; [7]Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [8]Division of Biostatistics, Department of Population Health, New York University Grossman School of Medicine, New York, NY, USA; [9]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; [10]Microsoft Research, Redmond, WA, USA; [11]Global Biometrics and Data Management, Pfizer Inc., New York, NY, USA; [12]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA; [13]Syneos Health Clinical Solutions, Amsterdam, The Netherlands; [14]Syneos Health Clinical Solutions, Morrisville, NC, USA; [15]Novo Nordisk, Søborg, Denmark; [16]AdventHealth Translational Research Institute, Orlando, FL, USA; [17]Cooperative Studies Program Coordinating Center, VA Palo Alto Health Care System, Palo Alto, CA, USA; [18]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA; [19]Division of Imaging Diagnostics and Software Reliability, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD, USA; [20]Global Real World Evidence Group, Moderna, Cambridge, MA, USA; [21]National Center for Advancing Translational Sciences, Bethesda, MD, USA; [22]Graticule Inc., Newton, MA, USA; [23]Adaptic Health Inc., Palo Alto, CA, USA; [24]Novartis Pharma AG, Basel, Switzerland and [25]Division of Endocrinology, Department of Medicine, University of North Carolina, Chapel Hill, NC, USA



Define causality in regulatory context

"Central questions for drug development and licensing are to establish the existence, and to estimate the magnitude, of treatment effects: how the outcome of treatment compares to what would have happened to the same subjects under alternative treatment (i.e. had they not received the treatment, or had they received a different treatment). "

ICH E9 (R1) (2019) Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials https://bit.ly/2UBlREf

Causal Estimands

- Individual-level causal effect: $Y_i(1) - Yi(0)$
- Population-level causal effect
  = Average treatment effect (ATE)
  = $\mathrm{E}\{Y(1) - Y(0)\}$
- Subgroup-level causal effect
  = Average treatment effect among treated (ATT)
  = $\mathrm{E}\{Y(1) - Y(0)|T = 1\}$
- Can also define in terms of ratios, other sub-groups, etc.

Causal assumptions

- Consistency: $Y = T * Y(1) + (1 - T) * Y(0)$
  - $Yi = Yi(1)$ if subject i is treated (i.e., $Y_i = 1$)
  - $Yi = Yi(1)$ if subject i is control (i.e., $Y_i = 0$)
  - May not hold under poor treatment adherence, lost-to-follow-up, and interference
- No unmeasured confounding: $T \perp\!\!\!\perp \{Y(1), Y(0)\}|C$
  - Aka: Strong ignorability, conditional exchangeability, exogeneity, etc.
  - Clinical judgement is often required for regulatory evidence generation
- Positivity: $\Pr(T = t \mid C = c) > 0$ for all (t, c)

Limitations of observational studies

- No unmeasured confounding can be violated: $T \perp\!\!\!\perp \{Y(1), Y(0)\} \mid C$
- How about RCTs? $T \perp\!\!\!\perp \{Y(1), Y(0)\}$

Causal inference roadmap *

I. Describe the observed data and the data generating experiment/process

II. Specify a *realistic* model for the probability distribution of observed data

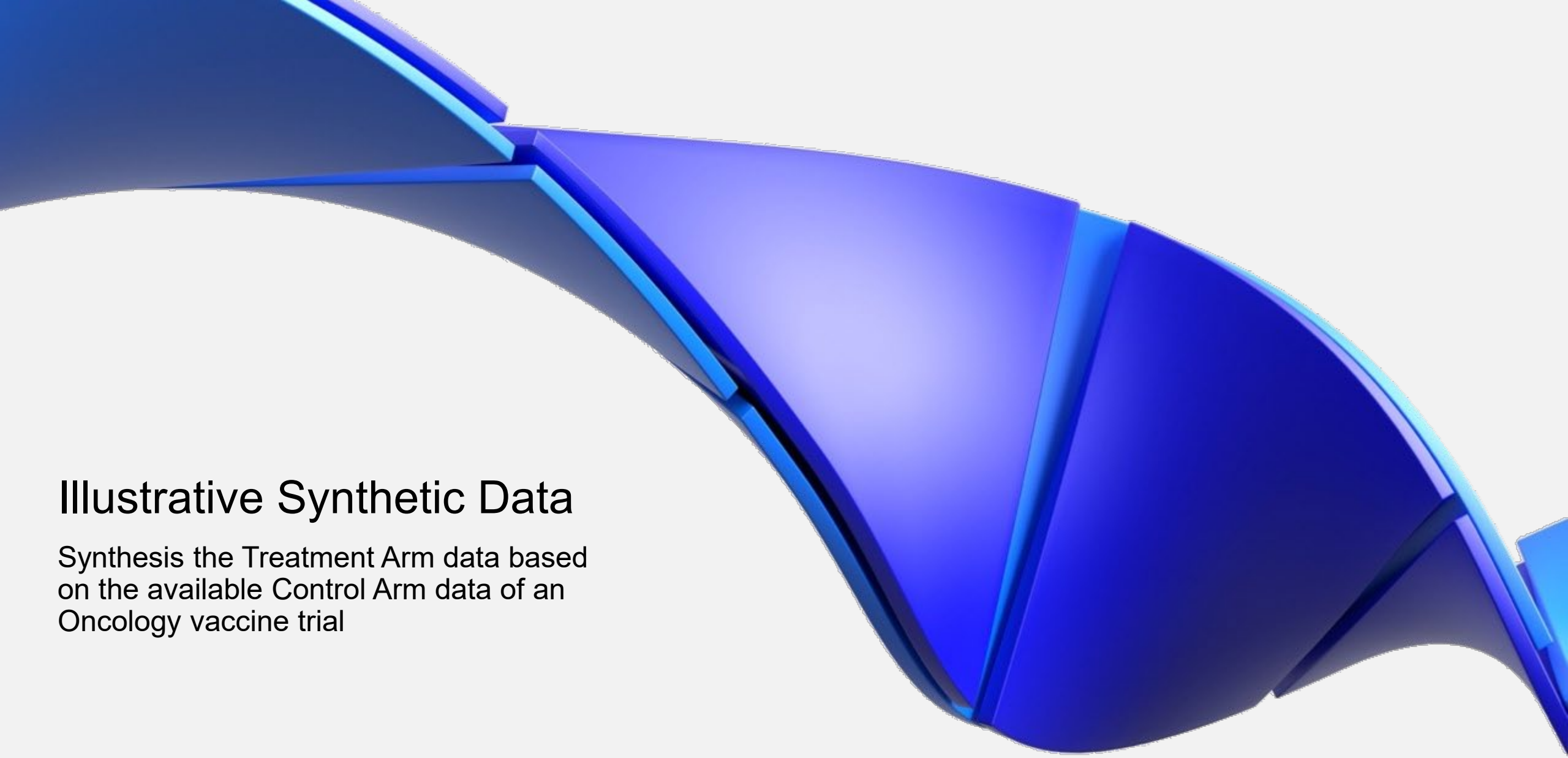
III. Define the causal and target estimand

IV. Select an estimator of the target estimand with desired properties

V. Develop an uncertainty estimator of the estimator for statistical inference

VI. Interpret the statistical results

* Ho et al. (2021) The Current Landscape in Biostatistics of Real-World Data and Evidence: Causal Inference Frameworks for Study Design and Analysis. DOI:10.1080/19466315.2021.1883475



Illustrative Synthetic Data

Synthesis the Treatment Arm data based on the available Control Arm data of an Oncology vaccine trial

Hypothetical examples inspired by a true trial…

- START trial (NCT00409188)
- Multi-center Phase III randomized, double-blind placebo-controlled
- Vaccine Stimuvax (L-BLP25 or BLP25 Liposome Vaccine)
- Control arm data only
- 504 subjects with unresectable stage III non-small cell lung cancer
- Each followed for 3 years
- Based on the control data to generate treatment arm data with desirable properties for the following illustrative examples

Sources: Project Data Sphere https://www.projectdatasphere.org/ ; https://clinicaltrials.gov/study/NCT00409188

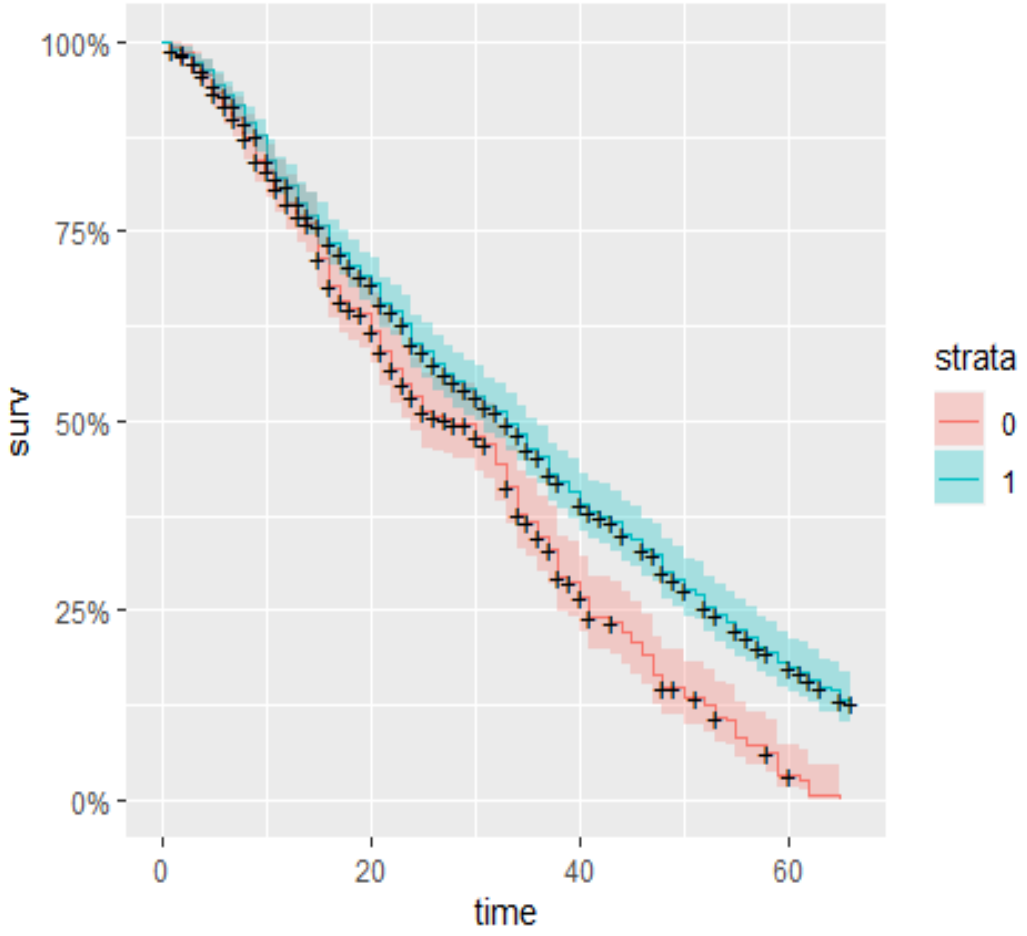
Baseline covariates of 500 control subjects

j	W_j	$W_j = 0$	$W_j = 1$	$W_j = 2$
1	Age	Mean 60.80 (SD=9.05) years		
2	Sex	32.0% female	68.0% male	--
3	Smoking	6.6% No	93.4% Yes	--
4	Histology	33.4% other	57.6% SCC	--
5	ECOG	56.8% Full Active	43.8% Restricted	--
6	Advanced stage	40% Stage IIIA	60% Stage IIIB or IIIC	--
7	Chemo. type	69.4% Concomitant	30.6% SCC	--
8	Chemo. response	69.6% Other	30.4% SCC	--
9	Region	37.4% Other;	41.0% West Europe	21.6% N Ameri.ca

SCC = Squamous Cell Carcinoma

Steps to create a synthetic vaccine subject cohort

1. Draw a patient with replacement from control arm
2. Create a vaccine patient who is more likely to be: Older, male, smokes, at Stage IIIA, and receiving squamous cell carcinoma chemotherapy, compared to the control patient
3. Create synthetic vaccine outcomes like the figure (blue = vaccine group vs. red = control group)
4. Combine Steps 2-3 outputs to synthesize a vaccine patient observation
5. Repeat Steps 1-4 by n_v times



Causal inference roadmap *

I. Describe the observed data and the data generating experiment/process

II. Specify a *realistic* model for the probability distribution of observed data

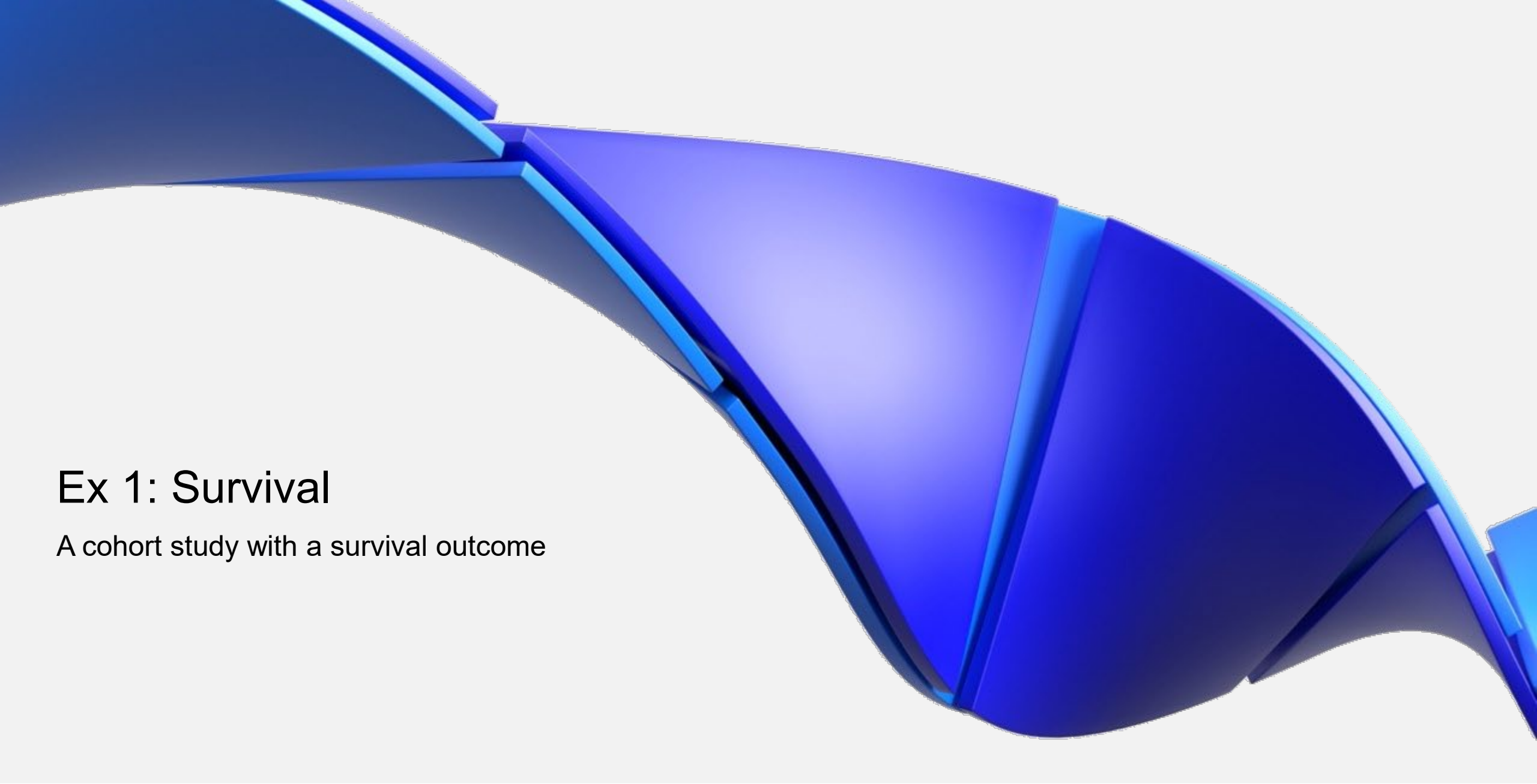
III. Define the causal and target estimand

IV. Select an estimator of the target estimand with desired properties

V. Develop an uncertainty estimator of the estimator for statistical inference

VI. Interpret the statistical results

* Ho et al. (2021) The Current Landscape in Biostatistics of Real-World Data and Evidence: Causal Inference Frameworks for Study Design and Analysis. DOI:10.1080/19466315.2021.1883475



Ex 1: Survival

A cohort study with a survival outcome

Step I: Describe observed data & data generating experiment

- $O = (A, W_1, \cdots, W_9, \Delta, T)$ = Observed data of a patient $\sim P_o$ $(i.i.d.)$
- $(W_1, \cdots, W_9)$ = Baseline covariates
- A = Assigned treatment (vaccine or not) at baseline
- Δ = Status, 1 (death) or 0 (censored)
- $T = \min(\tilde{T}, C)$ = Observed time, where $\tilde{T}$ = time to death and C = censoring time
- $N(t) = I(T \leq t, \Delta = 1)$ = Process of death
- $A(t) = I(T \leq t, \Delta = 0)$ = Process of censoring
- n_c = 500 and n_v = 1,000

Step II: Specify a realistic model for observed data distribution (1)

- Observed longitudinal data of a patient: $O = \{W, A, (N(t), A(t): t = 1, \cdots, K)\}$
- Structural causal model:

$$W = f_W(U_W)$$

$$A = f_A(W, U_A)$$

$$dN(t) = f_{dN(t)}(P_a\big(dN(t), U_{dN(t)}\big), t = 1, \cdots, K$$

$$dA(t) = f_{dA(t)}(P_a\big(dA(t), U_{dA(t)}\big), t = 1, \cdots, K$$

where $P_a\big(dN(t)\big) = \big(W, A, \bar{N}(t-1), \bar{A}(t-1)\big)$ denotes the history before $dN(t)$ is realized &

$P_a\big(dA(t)\big) = \big(W, A, \bar{N}(t), \bar{A}(t-1)\big)$ denotes the history before $dA(t)$ is realized.

Step II: Specify a realistic model for observed data distribution (2)

- Assuming all exogeneous variables $\boldsymbol{U} = \left(U_A, U_W, U_{dN(t)}, U_{dA(t)}\right)$ are independent → sequential independent assumption is hold for both death and censoring indicators.
- Positivity assumption is also made for both survival and censoring functions.
- Some of these *causal assumptions* may not be testable.

Step III: Define the causal estimand and target estimand

- Define $T^a = T^{a,\bar{A}=\bar{0}}$ as the potential outcome of treatment $A = a$, we choose the average treatment effect (ATE) as the causal estimand:

$$\theta^* = P(T^1 > \tau) - P(T^0 > \tau), \text{ where } \tau = 12, 24, \text{ or } 36 \text{ months}$$

- Following casual assumptions in Step 2, the target estimand below is equal to ATE:

$$\theta = E_{W,0}[S_0(\tau|A = 1, W) - S_0(\tau|A = 0, W)],$$

where $S_0(\tau|A, W)$ is the conditional survival function of $\tilde{T}$ given A and W. And we can connect survival function with hazard function $\bar{Q}_0(t|A, W) = P_0(\tilde{T} = t|\tilde{T} \geq t, A, W)$,

$$S_0(\tau|A, W) = \prod_{t=1}^{\tau} \left(1 - \bar{Q}_0(t|A, W)\right).$$

Step IV: Select an estimator of the target estimand

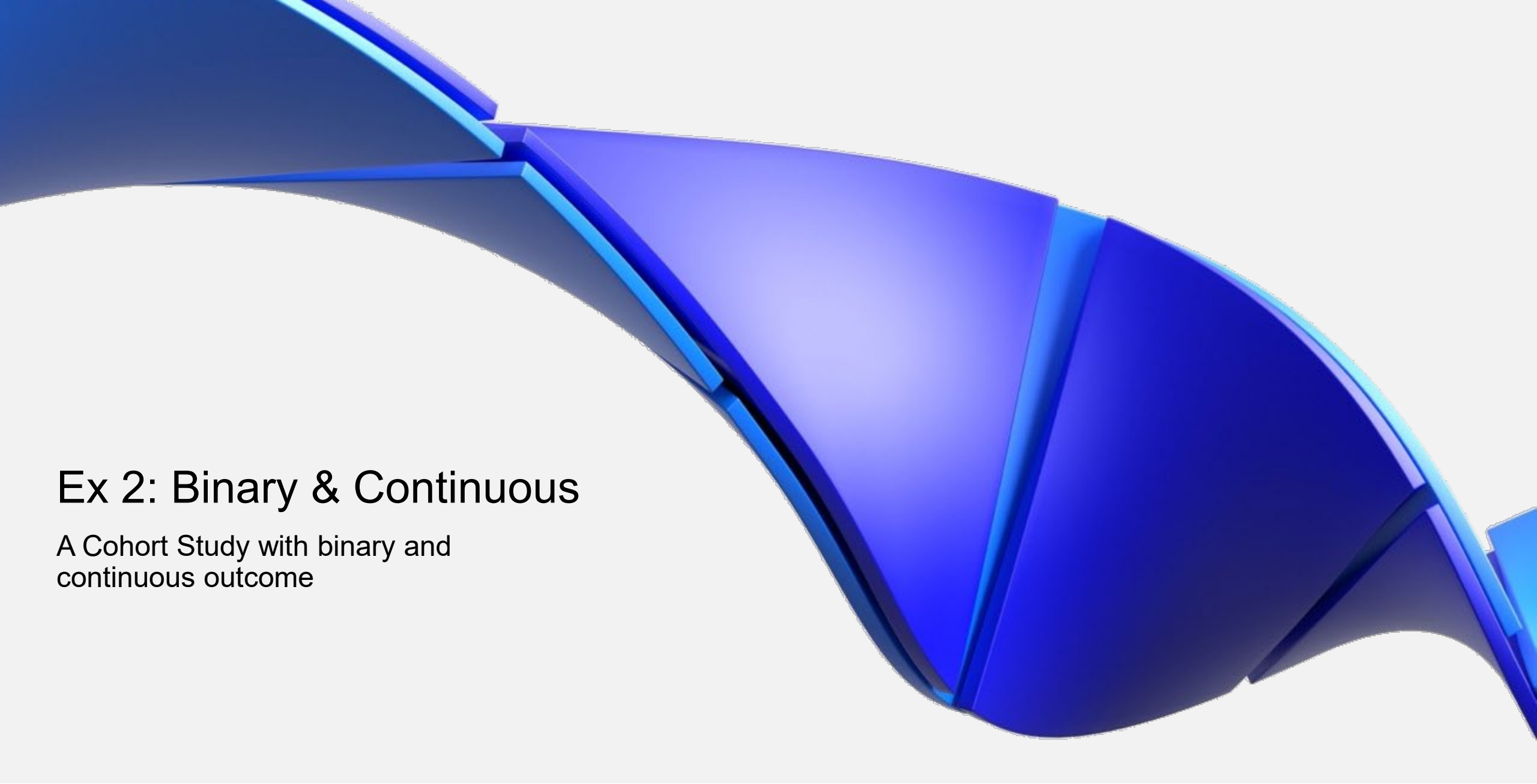
Step V: Develop an uncertainty estimator of the estimator

- Use the targeted maximum likelihood estimator or minimum loss-based estimator (TMLE) (van der Laan and Rose 2011)
- Asymptotically efficient under mild model assumptions
- R package "`suvtmle`" available to implement the TMLE approach
- The TMLE approach also provides an estimator of the uncertainty based on the efficient influence curve theory.

Step VI: Interpret the statistical results

- If the causal assumptions in Step II are plausible, the hypothetical vaccine's average treatment effect in terms of cumulative incidences of death by 12, 24, and 36 months are as follows:

τ	$\hat{\theta}$	SE	$95\%\ C.I.$
12 months	-1.1%	2.5%	$(-5.9\%, 3.7\%)$
24 months	-2.5%	3.8%	$(-9.8\%, 4.9\%)$
36 months	8.0%	4.1%	$(-0.1\%, 16.1\%)$



Ex 2: Binary & Continuous

A Cohort Study with binary and continuous outcome

Step I: Describe observed data & data generating experiment

- $O = (A, W_1, \cdots, W_9, Y) =$ Observed data of a patient $\sim P_o\ (i.i.d.)$
- $(W_1, \cdots, W_9) =$ Baseline covariates
- $A =$ Assigned treatment (vaccine or not) at baseline
- $n_c = 500$ and $n_v = 1{,}000$

Step I: Describe observed data & data generating experiment

- $O = (A, W_1, \cdots, W_9, Y) =$ Observed data of a patient $\sim P_o\ (i.i.d.)$
- $(W_1, \cdots, W_9) =$ Baseline covariates
- $A =$ Assigned treatment (vaccine or not) at baseline
- $n_c = 500$ and $n_v = 1{,}000$

Step II: Specify a realistic model for observed data distribution (1)

- Observed longitudinal data of a patient: $O = \{W, A, Y\}$
- Structural causal model:

$$W = f_W(U_W)$$

$$A = f_A(W, U_A)$$

$$Y = f_Y(W, A, U_Y)$$

Step II: Specify a realistic model for observed data distribution (2)

- Assuming all exogeneous variables $\boldsymbol{U} = \left(U_A, U_W, U_{dN(t)}, U_{dA(t)}\right)$ are independent → randomization assumption is hold:

$$Y^a \perp\!\!\!\perp A|W,$$

where $Y^a = f_Y(W, a, U_Y)$ is potential outcome for $a = 1{,}0$.

- Positivity assumption:

$$P(A = a|W = w) > 0, \text{for } a = 1{,}0 \text{ if } P_W(w) > 0.$$

Step III: Define the causal estimand and target estimand

- We select ATE as the causal estimand:

$$E_0(Y^1) - E_0(Y^0).$$

- Following casual assumptions in Step 2, the target estimand below is equal to ATE:

$$\Psi_{\mathrm{ATE}}(P_0) = E_{0,W}[E_0(Y|A=1,W) - E_0(Y|A=0,W)].$$

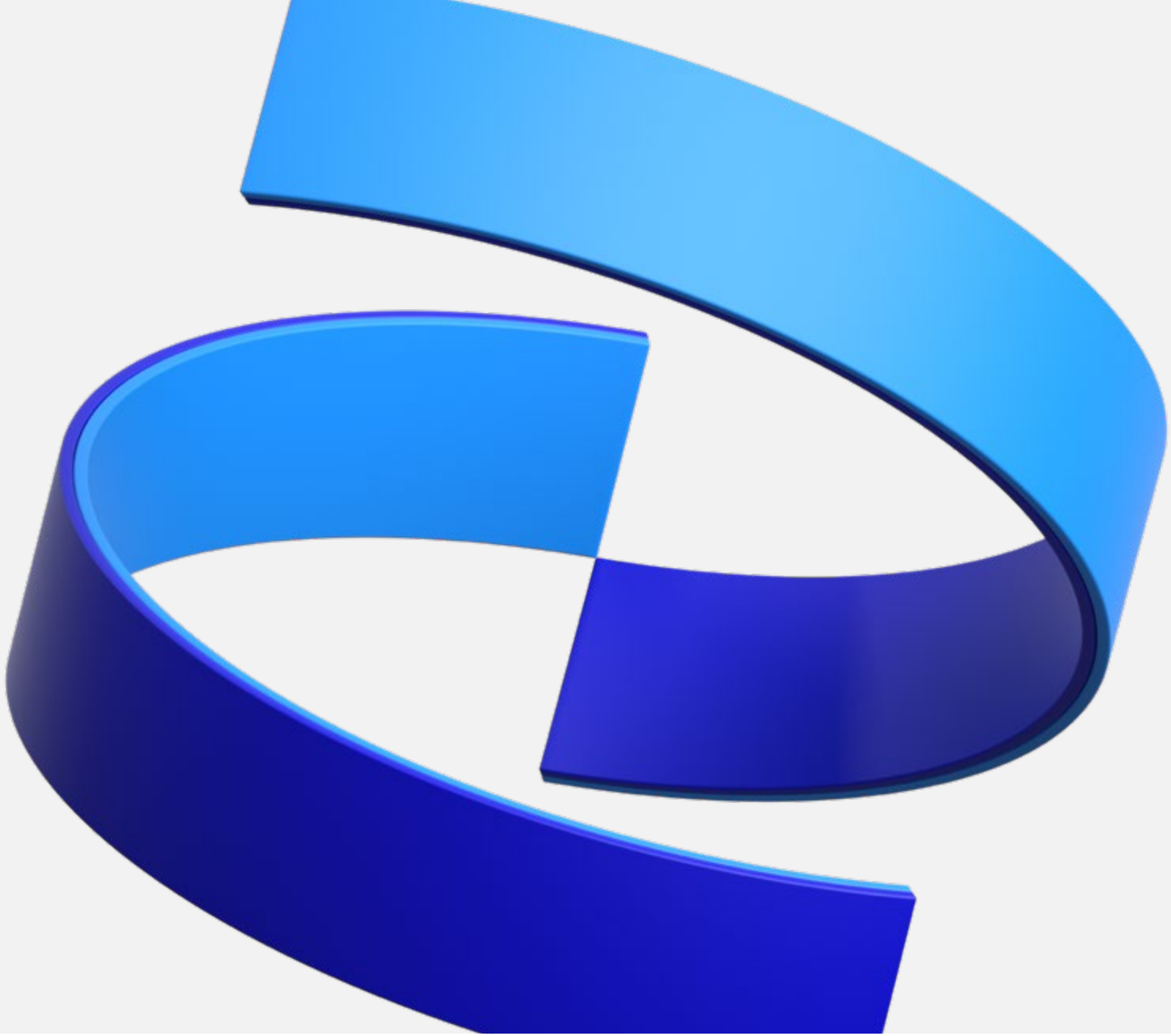
Step VI: Interpret the statistical results

- For binary outcome, the TMLE estimate $\widehat{\Psi}_{ATE} = 16.6\%$ with 95% $CI = (11.6\%, 21.7\%)$

Looking forward

- An example of a single arm study with an external control is in progress.
- Specifying the causal estimand of interest can guide statisticians in choosing the right target estimand in absence of randomization.
- The framework help statisticians decouple estimand and estimator and clarify the design and analysis of RWE studies.
- The framework make the causal assumptions explicit for casual interpretation of a RWE results.
- Average treatment effect of the treated can be useful.
- Various types of intercurrent events require further research.

Thank You



Backup Slides

Steps 1-2 to create vaccine arm data from control

1. Draw a control patient with replacement from the controlled arm of n_c patients $O = (A = 0, W_1, \cdots, W_9, \Delta, T)$, where $W_1, \cdots, W_9$ are baseline covariates,), Δ is status (death v censored), and $T = \min(\tilde{T}, C)$ is the observed time.

2. Create a vaccinated patient's baseline W_j' from the control patient's W_j, , $j = 1, \cdots, 9$:

$j = 1$	$W_1' = W_1 + N(\mu_1, 1)$
$j = 2, \ldots, 8$	$P(W_j' = 1 \mid W_j = 0) = p_{j,01}; P(W_j' = 0 \mid W_j = 1) = p_{j,10}$
$j = 9$	$P(W_9' = 1 \mid W_9 = 0) = p_{9,01}; P(W_9' = 2 \mid W_9 = 1) = p_{9,12};$ $P(W_9' = 0 \mid W_9 = 2) = p_{9,20}$

μ_1	$p_{j,01}$	$p_{j,10}$	$p_{8,01}$	$p_{8,10}$	$p_{9,01}$	$p_{9,12}$	$p_{9,20}$
3	0.1	0.2	0.1	0.3	0.1	0	0.2

Steps 3-5 to create vaccine arm data from control

3. Create the vaccine patient's outcomes (Δ', T') based on the control's outcome (Δ, T):

$P(\Delta' = 1 \mid \Delta = 0) = p_{\delta,01}$	$P(\Delta' = 0 \mid \Delta = 1) = p_{\delta,10}$
if $\Delta' = 1$	if $\Delta' = 0$
$T' = T \exp\{\beta_{\boldsymbol{t},0} + \sum_j \beta_{\boldsymbol{t},j} W_j + \mathrm{U}(-0.5, 0.5)\}$	$T' = T \exp\{\beta_{\boldsymbol{c},0} + \sum_j \beta_{\boldsymbol{c},j} W_j + \mathrm{U}(-0.5, 0.5)\}$

$p_{\delta,01}$	$p_{\delta,10}$	$(\beta_{t,0}, \cdots, \beta_{t,9})$	$(\beta_{c,0}, \cdots, \beta_{c,9})$
0.1	0.2	$(-6, 0.1, 0, -0.1, 0.1, 0, 0.3, -0.3, 0.2, 0)/3$	$(0, \cdots, 0)$

4. Combine the values from Steps 2-3 into an observation of a vaccinated patient
5. Repeat Steps 1-4 by n_v times -> vaccinated patients with covariates and outcomes.