Time-Varying Treatments in Observational Studies: Lessons for Clinical Trials

Vanessa Didelez
Leibniz Institute for Prevention Research and Epidemiology – BIPS
and Dpt. Mathematics and Comp.Science, University of Bremen

BBS – September 2020

Overview

- Estimands
- Observational studies: assumptions and methods
  - Treatment rules
  - Endpoints
- Lessons for RCTs
- Conclusions

Preliminaries

- I am a statistician working on theory and methods for the analysis of observational data in epidemiology
- I have no particular expertise in clinical trials
- ICH E9 Addendum:
  - Makes many good and important points
  - Some unfortunate choices of terminology ignoring much existing literature
  - Some important omissions

Estimands?

- What is the aim of our analysis / what is our research question? Target of inference?

- “Estimand” = what we want to estimate

  - Also (in statistical theory): what an estimator estimates when assumptions (not) satisfied

  ⇒ May or may not be a parameter parameterising (a part of) our model

Estimands – in RCTs

- In RCTs, need to decide on estimand especially with view to “intercurrent events”:
  - alternative treatments / rescue medication
  - discontinuation of treatment
  - switching
  - dropout
  - terminal events / competing risks

⇒ occurrence after randomisation – source of bias

Causal Estimand

= "treatment rules" to be compared

(individual interventions, combinations of interventions, regimen of sequence of interventions)

aka: "treatment strategies"

+ target population / subgroups

+ well defined **outcome(s) / endpoint(s)**

+ desired statistical contrast

(e.g. survival probs. vs HR)

Observational Studies

Early HIV studies (1990/2000s) (Swiss HIV Cohort Study!)

- No RCTs available, but treatment decisions required
- “Effect” of ART on time to AIDS / survival?
- Varying start of treatment (CD4 / viral load)
- Issues
  - Severe side-effects, non-adherence
  - Per protocol: “always” versus “never” treat?
  - When best to start?
  - When best to switch? (in view of failure / resistance)
- Main motivation for marginal structural models

Hernan et al (2000); Cain et al (2016); Gran et al (2010)

Observational Studies

Pharmaco-epidemiology (from health records / claims data)

- Post-accreditation safety / comparative effectiveness
- Wanted: long-term (side-)effect in general population (incl. vulnerable subgroups)
- Issues
  - Confounding by indication
  - Alignment of eligibility, treatment start, start of follow-up ambiguous
  - Combination of several medications, switching very common
  - Immortal time / prevalent user bias & other “self inflicted” biases
- Key example for target trial emulation

Hernan & Robins (2016), Labreque & Swanson (2017)

Time-Dependent Exposures / Treatments

Key characteristic:

- Exposure (treatment) is rarely a binary point-treatment; instead it is time-dependent (dynamic)
  - treatment / exposure often has a duration
  - can start / stop / start again / switch / combine etc.
  - often: want adaptive treatments

- Experience:
  trying to “force” this into simple binary point-treatment framework usually results in bias or practically irrelevant results

Time-Dependent Exposures / Treatments

- Need to be clear what treatment rules (sequence of interventions) we want to contrast
  - Relevant duration “less than…” vs. “more than…”
  - “Always treat” versus “never treat”
  - “If… then…” rules
  - “Start treatment immediately” vs. “delay treatment until…”

- To obtain inference on the effect of such treatment rules from observational data, need
  - defendable assumptions
  - suitable methods

How Do We Know What We Want?

Two (complementary) approaches:

Patient, doctor
Public health authority
(Regulatory agency?)

1. What is the **decision problem** you want to solve?

(Dawid & Didelez, 2010; Dawid, 2015)

- aim: compare the options between which to decide

2. What is the ideal experiment, **"target trial"**?

(Hernan & Robins, 2016; Cain et al, 2016)

- disregarding practical, ethical, financial, legal etc. constraints
- but *in this world & respecting physical laws*

⇒ then emulate trial as closely as possible with available data

How Do We Formalise What We Want?

Notation: to distinguish intervention from observation

Mathematically (here for **point** treatment X, endpoint Y):

- potential outcomes $Y(x)$
  - the value of Y if X were set to x by the intervention"
- **do-notation** $P(Y; do(X=x))$ & causal DAGs *(Pearl, 2009)*
  - "the distribution of Y if X were set to x by the intervention"
- RCTs with randomised X & full compliance: directly observe $P(Y(1))$ (treatment arm) and $P(Y(0))$ (controls)
- Note: we never observe $Y(1)$ and $Y(0)$ jointly princ.stratum be aware (and beware) of assumptions on joint distribution - *fundamentally untestable*. (Dawid, 2000)

Formalising Treatment Rules

Consider treatments over two seq. time points: X_1, X_2

Treatment rule

= rule for assigning values x_1, x_2 by seq. interventions

- Fixed in advance: "always $(x_1=1, x_2=1)$ / never treat $(x_1=0, x_2=0)$" or "stop early" $(x_1=1, x_2=0)$
- Or **dynamic / adaptive /…** :
  after X_1, observe $Z=z$, then assign $x_2=g(z)$ "g-methods"
  - Z could be CD4 count, side effect….

Formalising Treatment Rules

- Estimand: contrasts of different treatment rules
  - taking intercurrent events into account

- Example: compare

  1. x_1 = new drug, Z = occurrence of adverse event,
     x_2 = switch to std. drug if $Z=1$ else, x_2 = new drug

  with

  2. x_1 = std. drug, x_2 = std. drug

Identifiability Assumptions

Not mentioned in ICH E9 Addendum:

Under what structural assumptions can we identify our chosen estimand from observable data?

- in principle, N=∞
- non-parametrically

- In RCT: only treatment *assignment* is randomised
  - identfies ITT-based effects

Other assumptions (required by most approaches):

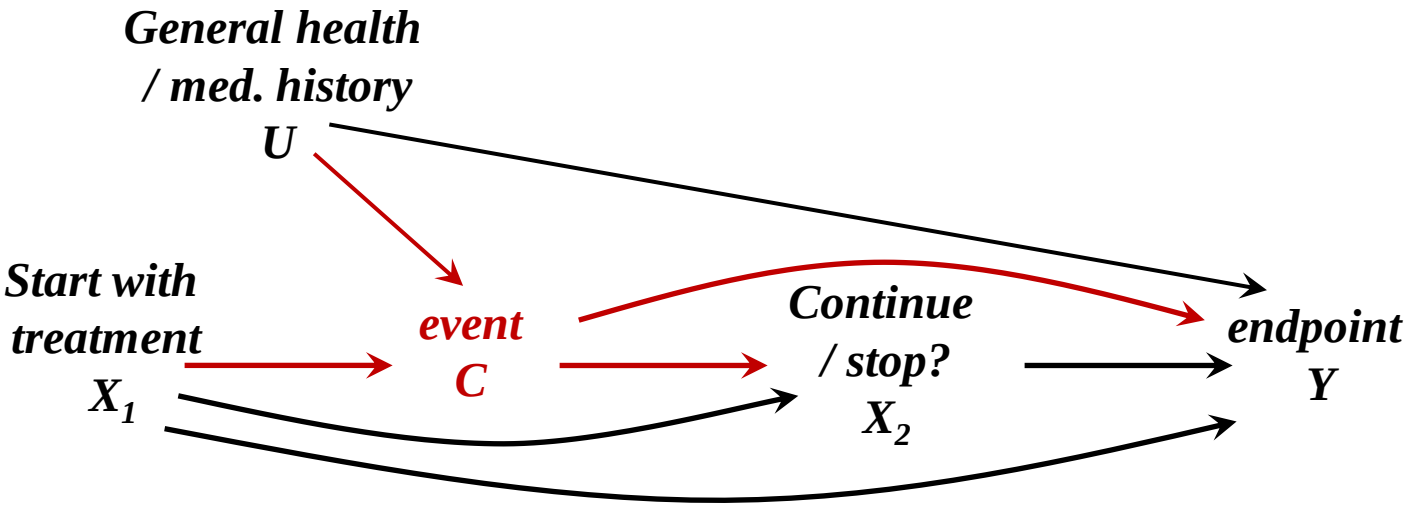
- Consistency → treatment rule "well-defined"
- Positivity → target treatment rules possible for all individuals
- No unmeasured time-varying confounding
  - aka "sequential randomisation"

Positivity Assumption

- Violated if chosen treatment rules cannot possibly occur for certain patients, e.g. “continue with treatment even under allergic reaction”
  - if allergic reaction is predictive of endpoint
- If violated, practical relevance of estimand debatable
  - alluded to in ICH E9 Addendum
- Mathematically:
  - we would extrapolate into area where we have no data – may not notice under parametric models
  - with “inverse-probability-weighting”: would divide by zero

Time-Dependent Confounding

(Robins, 1986+)



- C affected by prior treatment, confounding future treatment and endpoint
- Note:
  - $\boldsymbol{U}$ can remain unmeasured if C is measured
    = “sequential randomisation” (cond. on C)
  - $\boldsymbol{U}$ and $\boldsymbol{X_1}$ independent by randomisation

Adjusting for Time-Dep. Confounding

- Cannot use:
  - Regression adjustment
  - (Propensity score) matching / stratification

Instead:

- g-formula
- MSMs fitted by IPW
  - time-varying weights

basic principles

⇒ Basis for many more advanced techniques

- Double robust estimation / g-estimation
- Targeted maximum likelihood
- … combine with machine learning approaches

g-Formula

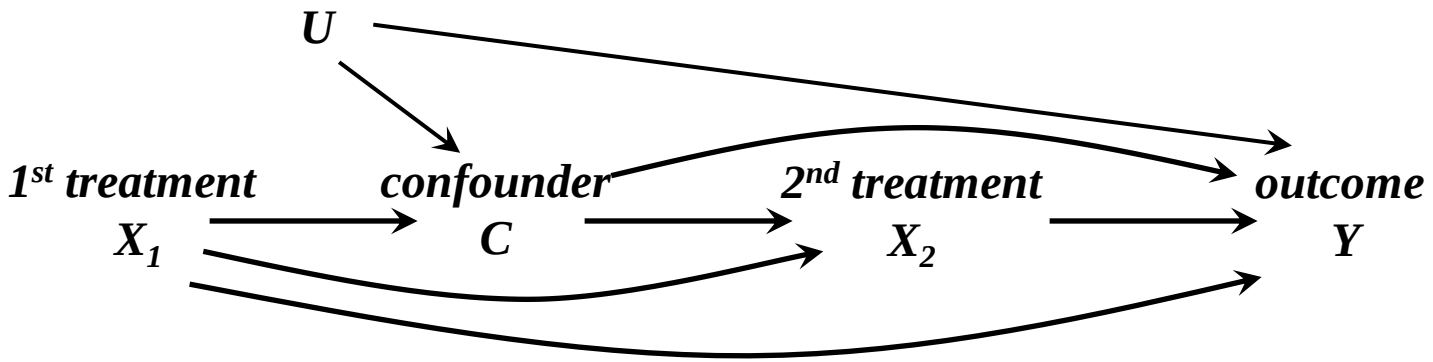
(Robins, 1986; reformulated: Dawid & Didelez, 2010)

- Need models for
  - $E(Y \mid X_1 = x_1,\ C{=}c,\ X_2 = x_2)$
  - $P(C \mid X_1 = x_1)$

For $(X_1 = x_1, X_2 = x_2)$ as determined by treatment rule:

$$\sum_c E(Y \mid X_1 = x_1,\ C{=}c,\ X_2 = x_2)\, P(C \mid X_1 = x_1)$$

(early application: Robins, Hernan, Siebert, 2004)



Inverse Probability of Treatment Weighting

- Assume marginal structural model (MSM)
  - for $E(Y;$ *treatment rule A*$)$ vs. $E(Y;$ *treatment rule B*$)$
  - or for dynamic strategy $E(Y; g(.))$
- Fit by weighting each obs. with inverse of prob. for observed treatment sequence

$$P(X_1 = x_1)\, P(X_2 = x_2 | C{=}c)$$

  - re-weighted population: like randomised
- Special cases
  - Survival outcome – time-varying weights;
  - "Optimal" strategy – regret regression models, optimal MSMs

Murphy (2003), Chakraborty & Moodie (2013)

“Quality of Life after Death”

- Intercurrent event: terminal (comp. risk)
- Example: want effect of treatment on cognitive function in the elderly
  - endpoint not observable if death occurs first
- Patient: might want to know whether treatment affects survival…
  - arguably, more important than cognitive function
- ICH E9: consider “composite”, i.e. combined endpoints
  - lessons from observational studies…?

“Quality of Life after Death”

Sensible estimand?

- Not: subgroup = “alive” (selection bias)
  - except if death known to be unaffected by treatment!
- Practically relevant: combine death & cognitive function
  - Utility function?
  - Note: joint distribution after randomisation is identified
- *Principal stratum* subgroup[illegible]rs”
  arguably not directly useful for decision making nor implementable in target trial *(Dawid & Didelez, 2011)*

relevant subgroup?

- New approach:
  separable effects *(Didelez, 2019, Stensrud et al, 2020)*

Conclusions

Lessons

- There is no such thing as “a / the treatment effect”
- There are only contrasts of specified *treatment rules*
  - Most treatments in practice get changes / adapted over time
  - We should target treatment rules with practical relevance
  - Taking possible changes over time into account

Implications for RCTs

Design trumps analysis

- When you *can* actually carry out your target trial, then do it!
  - Randomise between (dynamic / adaptive) treatment rules relevant for real practice, including possibilities of dealing with intercurrent events
  - … this will also increase adherence Scharfstein (2019)
  - e.g. “treat with A except if Z happens, then switch to B” etc. if this best reflects real life decisions

Implications for RCTs

- Statistical methods to compensate for (time-varying) confounding, or other “statistical tricks”, should only be used as last resort

- Collect enough information on reasons for intercurrent events
  - to plausibly assume “no unmeasured time-varying confounding”
  - especially: to characterise adherence

Advert 1 – ofcaus.org

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Topic Group 7

https://www.ofcaus.org/home 67%

STRATOS INITIATIVE Topic Group 7

Home About Us Causal inference literature Courses Estimation approaches and data e... R implementations

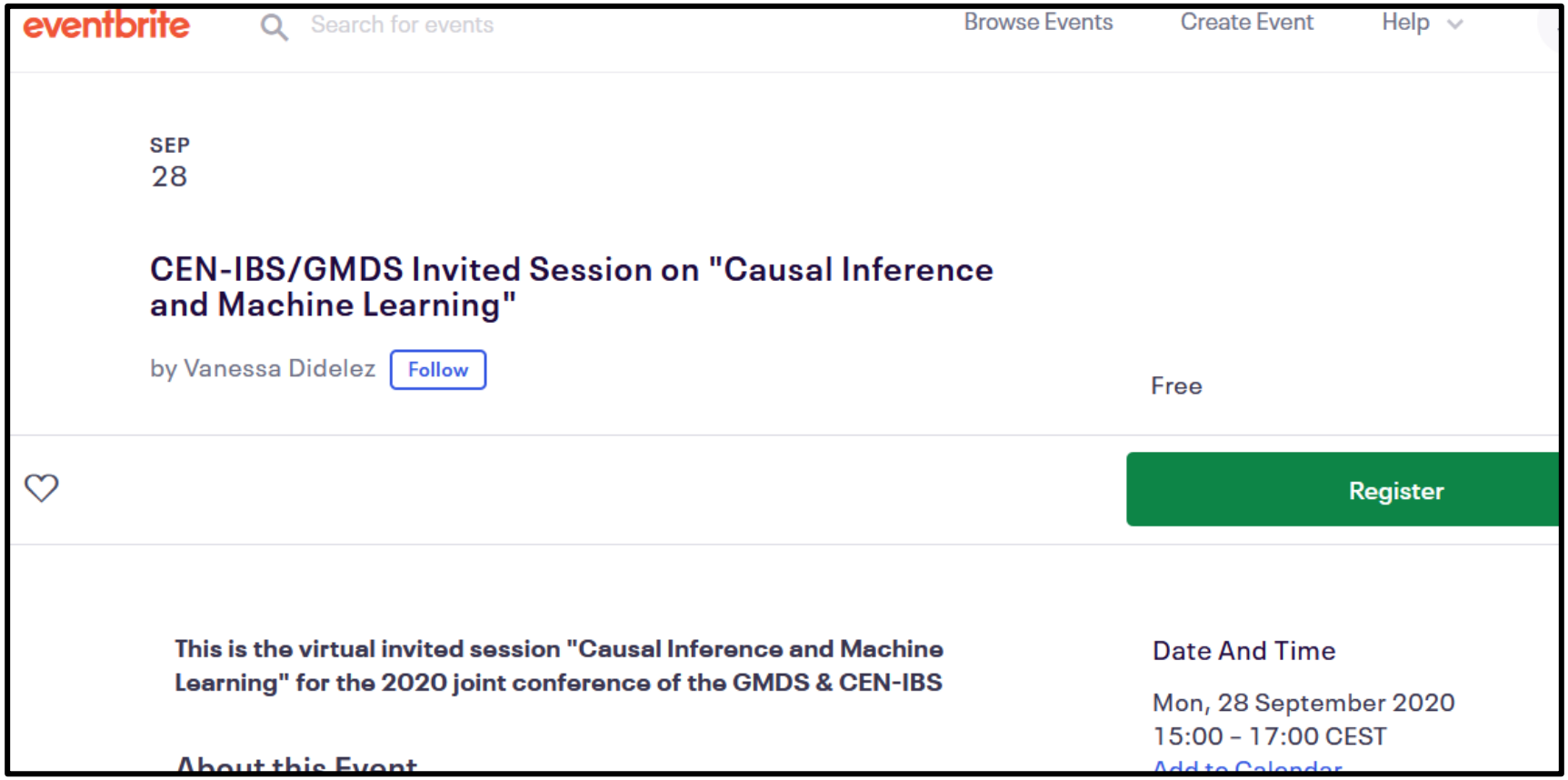
Causal Inference

STRATOS Initiative

Topic group 7 is a member of the STRATOS Initiative (STRengthening Analytical Thinking for Observational Studies) which is a large collaboration of experts in many different areas of biostatistical research. Ongoing research, discussions and activities within STRATOS are conducted in nine topic groups and several cross-cutting panels.

Advert 2 – Invited Session

- https://www.eventbrite.de/e/cen-ibsgmds-invited-session-on-causal-inference-and-machine-learning-tickets-116222778459



Thanks!

www.leibniz-bips.de/en

Contact

Vanessa Didelez

Leibniz Institute for Prevention Research
and Epidemiology – BIPS
Achterstraße 30
D-28359 Bremen

References

- Cain, Saag, Petersen et al (2016). Using observational data to emulate a randomized trial of dynamic treatment switching strategies: an application to antiretroviral therapy. *IJE.*
- Cartwright (2003). Nature's Capacities and Their Measurement. *Oxford Scholarship Online*
- Chakraborty, Moodie (2013). Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference and Personalized Medicine. *Book*
- Dawid (2015). Statistical Causality from a Decision-Theoretic perspective. *Annual Review of Statistics and Its Application*
- Dawid, Didelez (2010). Identifying the consequences of dynamic treatment strategies: A decision theoretic overview. *Statistics Surveys*
- Didelez (2016). Commentary: Should the analysis of observational data always be preceded by specifying a target experimental trial? *Oxford University Press*
- Didelez (2019). Defining causal mediation with a longitudinal mediator and a survival outcome. *LIDA*
- Fox, Edwards, Platt, Balzer (2019). The critical importance of asking good questions: The role of epidemiology doctoral training programs. *American Journal of Epidemiology*
- Gran, J.M., Røysland, K., Wolbers, M., Didelez, V., Sterne, J.A.C., Ledergerber, B., Furrer, H., von Wyl, V. and Aalen, O.O. (2010), A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. Statist. Med., 29: 2757-2768.
- Goetghebeur, le Cessie, De Stavola, Moodie, Waernbaum (2019). Formulating causal questions and principled statistical answers. *To appear in Stat in Med.*
- Hernán (2016). Does water kill? A call for less casual causal inferences. *Annals of Epidemiology*

References

- Hernán, Alonso, Logan (2008). Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*
- Hernán, M., Brumback, B., & Robins, J. (2000). Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology, 11(5), 561-570.
- Hernán, Robins (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology*
- Labrecque, Swanson (2017). Target trial emulation: teaching epidemiology and beyond. *European Journal of Epidemiology*
- Lash, Fox , Fink (2011). Applying quantitative bias analysis to epidemiologic data. *Springer Science & Business Media*
- Murphy (2003) . Optimal dynamic treatment regimes. *JRSSB*
- Pearl (2000, 2009). Causality – models, reasoning and inference. *Campridge University Press*
- Robins (1986). A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker effect. *Mathematical Modelling*
- Robins, Hernán, Siebert (2004). Effects of multiple interventions. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*
- Scharfstein DO. A constructive critique of the draft ICH E9 Addendum. Clin Trials. 2019;16(4):375-380.
- Stensrud, Young, Didelez, Robins, Hernan (2019): Separable effects for causal inference in the presence of competing risks. *arXiv*

References on Principal Stratification

- Pearl J. (2011). Principal stratification--a goal or a tool?. *The International Journal of Biostatistics*, *7*(1), 20.

And discussions of the above paper :

- Dawid, P., & Didelez, V. (2012). "Imagine a Can Opener"--The Magic of Principal Stratum Analysis, *The International Journal of Biostatistics*, *8*(1).
- Gilbert, P. B., Hudgens, M. G., & Wolfson, J. (2011). Commentary on "Principal stratification - a goal or a tool?" by Judea Pearl. *The international journal of biostatistics*, *7*(1), 36.
- Joffe M. (2011). Principal stratification and attribution prohibition: good ideas taken too far. *The international journal of biostatistics*, *7*(1), 35.
- Mealli, F., & Mattei, A. (2012). A Refreshing Account of Principal Stratification, *The International Journal of Biostatistics*, *8*(1).
- Vanderweele T. J. (2011). Principal stratification--uses and limitations. *The international journal of biostatistics*, *7*(1), 28.
- *and find many more discussions in that journal*