Day 3, Lecture 1

Longitudinal data structures and time-dependent treatment decisions

Overview of course: Day 3

Day 3: 8 – 10

Time-dependent treatment decisions. Causal inference in longitudinal data.

- Treatment-confounder feedback.

Day 3: 10 – 12

Longitudinal TMLE. Targeting for time-varying structures.

- Identification proofs and extension of the time-fixed setting.
- Software: `ltmle`.

Lunch.

Day 3: 13 – 15

Evaluation + "buffer".

Longitudinal data structures

Lecture 1 What are we targeting?

- Time-varying treatment interventions.
- Identification and time-dependent confounding.
- An introduction to get started with Lecture 2.

Lecture 2 TMLE for estimation

- IP-weighting + G-formula.
- Iterated expectations representation.
- Targeting effects of time-varying treatment interventions.
- `ltmle` software package

Hernán et al., 2008

Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease

Miguel A. Hernán[1,2], **Alvaro Alonso**[3], **Roger Logan**[1], **Francine Grodstein**[1,4], **Karin B. Michels**[1,4,5], **Meir J. Stampfer**[1,4,6], **Walter C. Willett**[1,4,6], **JoAnn E. Manson**[1,4,7], and **James M. Robins**[1,8]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, MA

[2]Harvard-MIT Division of Health Sciences and Technology, University of Minnesota, Minneapolis, MN

[3]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN

[4]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[5]Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

[6]Department of Nutrition, Harvard School of Public Health, Boston, MA

[7]Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

[8]Department of Biostatistics, Harvard School of Public Health, Boston, MA

Abstract

Background—The Women's Health Initiative randomized trial found greater coronary heart disease (CHD) risk in women assigned to estrogen/progestin therapy than in those assigned to placebo. Observational studies had previously suggested reduced CHD risk in hormone users.

Methods—Using data from the observational Nurses' Health Study, we emulated the design and intention-to-treat (ITT) analysis of the randomized trial. The observational study was conceptualized as a sequence of "trials" in which eligible women were classified as initiators or noninitiators of estrogen/progestin therapy.

Hernán et al., 2008 (part 1/2)

An example of the importance of being clear about what effect we are targeting...

(and the general difficulties in analyzing observational studies).

Hernán et al., 2008 (part 1/2)

An example of the importance of being clear about what effect we are targeting...

(and the general difficulties in analyzing observational studies).

Background for the paper:

- An RCT found **greater** risk of coronary heart disease (CHD) in women assigned to hormon therapy than those assigned to placebo.
- Earlier observation studies had found **reduced** risk of CHD among hormon users.

The difference has been explained as due to unobserved confounding.

Conclusion: *Cannot use observational data for causal inference?*

Hernán et al., 2008 (part 1/2)

The **RCT results** were based on an intention-to-treat (ITT)[1] analysis

- trial participants were randomized to hormone treatment initiation or placebo at baseline

[1] Subjects are analyzed irrespective of their actual adherence to their assigned randomization arm.

Hernán et al., 2008 (part 1/2)

The **RCT results** were based on an intention-to-treat (ITT)[1] analysis

- trial participants were randomized to hormone treatment initiation or placebo at baseline

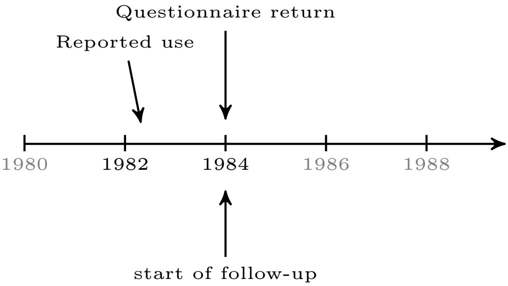
Whereas the **observational analyses** were based on a comparison of two groups:

- "Current users"
- "Never users"

[1] Subjects are analyzed irrespective of their actual adherence to their assigned randomization arm.

Hernán et al., 2008 (part 1/2)

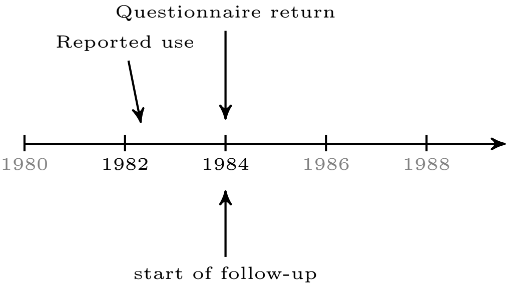
- In the observational study, women answered questionnaires every two years
  - updated information on use, duration, etc, of treatment



- The start of follow-up was defined as the return of the questionnaire
  - initiaters who stopped/died before return were excluded (to define "current-users")

Hernán et al., 2008 (part 1/2)

- In the observational study, women answered questionnaires every two years
  - updated information on use, duration, etc, of treatment



- The start of follow-up was defined as the return of the questionnaire
  - initiaters who stopped/died before return were excluded (to define "current-users")

Selection bias: Early (harmful) effect of treatment not identified.

Hernán et al., 2008 (part 1/2)

Hernán et al. reanalyze the observational data and show in their paper that:

1. When using the current user design (including the selection bias), the result of a **beneficial** effect from earlier observational studies was reproduced.
2. *When imitating the analysis of the randomized trial*, targeting the ITT effect, the result that the treatment has a **harmful** effect was reproduced.

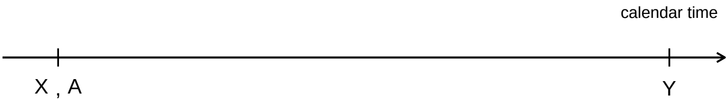
The discrepancy found in the previous analyses had nothing to do with confounding.

Causal inference for longitudinal data structures

New setting: Longitudinal data structure

Data structure considered so far:

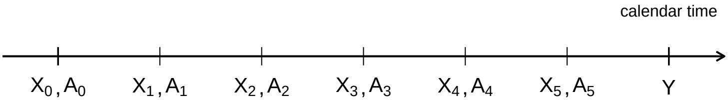
- $O = (X, A, Y) \in \mathbb{R}^d \times \{0,1\} \times \{0,1\}$
- Covariates X are measured before treatment decision A is made
- After treatment decision A, the outcome Y is observed



New setting: Longitudinal data structure

Longitudinal data structure:

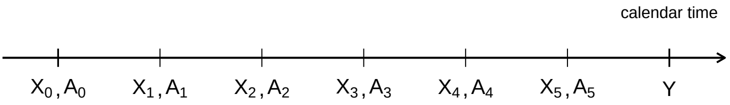
- $O = (X_0, A_0, X_1, A_1, \dots, X_K, A_K, Y = X_{K+1}) \in (\mathbb{R}^d \times \{0,1\})^K \times \{0,1\}$
- Covariates $X = (X_0, X_1, \dots, X_K)$ change over time
- Treatment decisions $A = (A_0, A_1, \dots, A_K)$ are updated over time
- Covariates and treatment decisions interact in complex ways



New setting: Longitudinal data structure

Longitudinal data structure:

- $O = (X_0, A_0, X_1, A_1, \ldots, X_K, A_K, Y = X_{K+1}) \in (\mathbb{R}^d \times \{0,1\})^K \times \{0,1\}$
- Covariates $X = (X_0, X_1, \ldots, X_K)$ change over time
- Treatment decisions $A = (A_0, A_1, \ldots, A_K)$ are updated over time
- Covariates and treatment decisions interact in complex ways



1. More complex treatment interventions.
2. More subtleties in confounding bias.
3. Right-censoring and competing risks.

$\vdots$

Longitudinal data structure

This data structure matches quite well the data collected in a **randomized clinical trial** with follow-up visits:

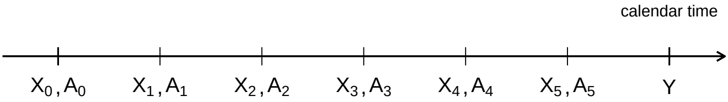
- X_0 are baseline covariates (age, sex, disease/medical history, ...)
- A_0 tells us the randomization arm (treatment/placebo)

  ⋮

- X_k are covariates measured at the kth follow-up visit
- A_k is the treatment decision made at the kth follow-up visit (adherence to randomization arm)

  ⋮

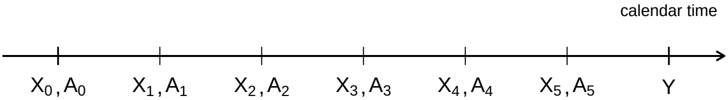
- Final outcome Y



Longitudinal data structure (sidenote #1)

When data is not randomized (i.e., **observational**).

- the time-grid data structure may be a bit artificial;[2]
- but otherwise the difference mostly consists in the randomized treatment decision at baseline.



[2] There may be more data modeling choices to make it fit nice structure.

Longitudinal data structure (sidenote #2)

Over the course of time of a study, we may not be able to observe the outcome of interest due to:

Loss to follow-up (right-censoring) For some individuals the event of interest is not known.

Presence of competing risk events No one cannot get experience the outcome event of interest if they already died.

How we handle these complications is reflected in our

- formulation of causal parameter;
- formulation of ideal interventions.

Longitudinal data structure (sidenote #2)

Example: Trial comparing treatment versus placebo on survival chances.

Causal question *What is the effect of the treatment?*

Longitudinal data structure (sidenote #2)

Example: Trial comparing treatment versus placebo on survival chances.

Causal question *What is the effect of the treatment?*

- Patients experiencing deteriorating health conditions are more prone to drop out.
- Patients randomized to placebo experience worse health conditions.

Longitudinal data structure (sidenote #2)

Example: Trial comparing treatment versus placebo on survival chances.

Causal question *What is the effect of the treatment?*

- Patients experiencing deteriorating health conditions are more prone to drop out.
- Patients randomized to placebo experience worse health conditions.

$\Downarrow$

Causal question *What is the effect of the treatment had there been no loss to follow-up?*

- Intervention strategy: Treatment + prevent loss to follow-up.
- Control strategy: Placebo + prevent loss to follow-up.

Longitudinal data structure (sidenote #2)

Example: Trial comparing treatment versus placebo on discharge from the intensive care unit (ICU).

Causal question *What is the effect of the treatment?*

Longitudinal data structure (sidenote #2)

Example: Trial comparing treatment versus placebo on discharge from the intensive care unit (ICU).

Causal question *What is the effect of the treatment?*

- Patients experiencing deteriorating health conditions have lower survival changes.
- Patients randomized to placebo experience worse health conditions.

Longitudinal data structure (sidenote #2)

Example: Trial comparing treatment versus placebo on discharge from the intensive care unit (ICU).

Causal question *What is the effect of the treatment?*

- Patients experiencing deteriorating health conditions have lower survival changes.
- Patients randomized to placebo experience worse health conditions.

A hypothetical world where subjects cannot die is a weird world.

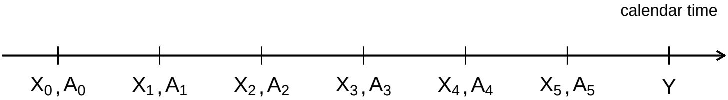
Longitudinal data structure (sidenote #2)

- Whenever the outcome of interest is not all-cause mortality, there can be competing risks.
- A competing risk event is not a right-censoring event.
- We are not (rarely?) interested in reporting the treatment effect *if subjects could not die*.

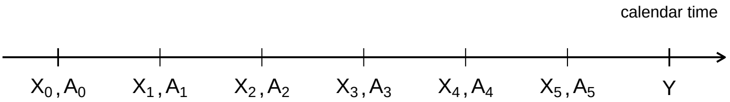
Longitudinal data structure

We consider (for now) the case that Y is fully observed:

- no right-censoring.
- no competing risks.



Time-dependent treatment interventions



What effect are we targeting?

Counterfactual outcomes

$$Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*}, \qquad \text{for,} \quad a_0^*, \ldots, a_K^* \in \{0, 1\}$$

= defined by a sequence of treatment decision rules that we choose.

also called:

- hypothetical treatment interventions
- hypothetical treatment strategies
- hypothetical treatment regimes

Time-dependent treatment interventions

Continued treatment and never treated:

- $Y^{A_0=1,A_1=1,\ldots,A_K=1}$ = outcome if treated throughout follow-up
- $Y^{A_0=0,A_0=0,\ldots,A_K=0}$ = outcome if untreated throughout follow-up

The risk difference

$$\mathbb{E}[Y^{A_0=1,A_1=1,\ldots,A_K=1}] - \mathbb{E}[Y^{A_0=0,A_0=0,\ldots,A_K=0}]$$

is the effect of being treated versus untreated throughout follow-up.

Time-dependent treatment interventions

Intention-to-treat (ITT):

- $Y^{A_0=1}$ = outcome if assigned to treatment arm
- $Y^{A_0=0}$ = outcome if assigned to placebo arm

The risk difference

$$\mathbb{E}[Y^{A_0=1}] - \mathbb{E}[Y^{A_0=0}]$$

is the effect of being assigned to the treatment versus the placebo arm

Hernán et al., 2008 (part 2/2)

Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease

Miguel A. Hernán[1,2], **Alvaro Alonso**[3], **Roger Logan**[1], **Francine Grodstein**[1,4], **Karin B. Michels**[1,4,5], **Meir J. Stampfer**[1,4,6], **Walter C. Willett**[1,4,6], **JoAnn E. Manson**[1,4,7], and **James M. Robins**[1,8]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, MA

[2]Harvard-MIT Division of Health Sciences and Technology, University of Minnesota, Minneapolis, MN

[3]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN

[4]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[5]Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

[6]Department of Nutrition, Harvard School of Public Health, Boston, MA

[7]Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

[8]Department of Biostatistics, Harvard School of Public Health, Boston, MA

Abstract

Background—The Women's Health Initiative randomized trial found greater coronary heart disease (CHD) risk in women assigned to estrogen/progestin therapy than in those assigned to placebo. Observational studies had previously suggested reduced CHD risk in hormone users.

Methods—Using data from the observational Nurses' Health Study, we emulated the design and intention-to-treat (ITT) analysis of the randomized trial. The observational study was conceptualized as a sequence of "trials" in which eligible women were classified as initiators or noninitiators of estrogen/progestin therapy.

Hernán et al., 2008 (part 2/2)

The current user strategy does not allow us to answer a causal question

1. The inclusion criterion is defined after initiation of the treatment strategy.
2. The current user strategy changes over the follow-up;
    - does not generally correspond to a counterfactual scenario with a sequence of treatment rules.

The current user strategy is an attempt to estimate the effect of treatment *usage* (contrary to initiation)?

Hernán et al., 2008 (part 2/2)

The current user strategy does not allow us to answer a causal question

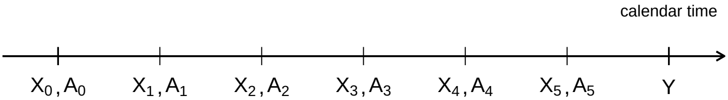
1. The inclusion criterion is defined after initiation of the treatment strategy.
2. The current user strategy changes over the follow-up;
    - does not generally correspond to a counterfactual scenario with a sequence of treatment rules.

The current user strategy is an attempt to estimate the effect of treatment *usage* (contrary to initiation)?

- Most appropriate summary measure would be the adherence-adjusted effect (comparing 'always treated' to 'never treated')?

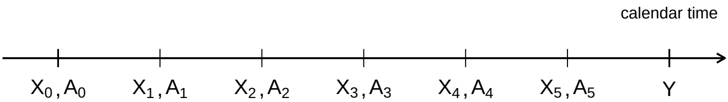
Hernán et al., 2008 (part 2/2)

Instead of the current user design, Hernán et al. divide the follow-up into monthly intervals



Hernán et al., 2008 (part 2/2)

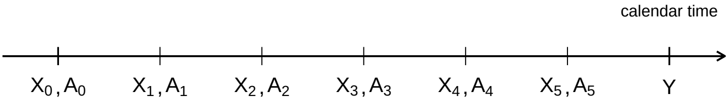
Instead of the current user design, Hernán et al. divide the follow-up into monthly intervals



- A_k = reported use of hormone therapy between month k and month $k+1$.

Hernán et al., 2008 (part 2/2)

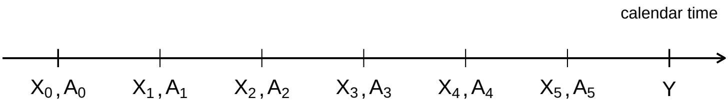
Instead of the current user design, Hernán et al. divide the follow-up into monthly intervals



- A_k = reported use of hormone therapy between month k and month $k+1$.
- ITT treatment strategies:
  1. Intervention strategy $A_0 = 1$
  2. Control strategy $A_0 = 0$

Hernán et al., 2008 (part 2/2)

Instead of the current user design, Hernán et al. divide the follow-up into monthly intervals



- A_k = reported use of hormone therapy between month k and month $k + 1$.
- ITT treatment strategies:
  1. Intervention strategy $A_0 = 1$
  2. Control strategy $A_0 = 0$
- Enforcing continued exposure (adherence-adjusted):
  1. Intervention strategy $A_0 = 1, A_1 = 1, A_2 = 1, \ldots$
  2. Control strategy $A_0 = 0, A_1 = 0, A_2 = 0, \ldots$

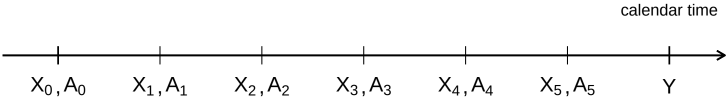
Example: Hernán et al., 2008

1. When using the current user design, the result of a **beneficial** effect from earlier observational studies (due to selection bias) was reproduced.
2. When imitating the analysis of the randomized trial, targeting the ITT effect, the result that the treatment has a **harmful** effect was reproduced.

Example: Hernán et al., 2008

1. When using the current user design, the result of a **beneficial** effect from earlier observational studies (due to selection bias) was reproduced.
2. When imitating the analysis of the randomized trial, targeting the ITT effect, the result that the treatment has a **harmful** effect was reproduced.
3. A **larger harmful** effect was found when targeting the effect of 'continued exposure', i.e., the adherence-adjusted effect.

Identification of effects of time-dependent treatment interventions



Time-dependent treatment interventions

Identification of $\mathbb{E}[Y^{A_0=a_0^*,A_1=a_1^*,\ldots,A_K=a_K^*}]$.

1. Consistency: $Y^{A_0=a_0^*,A_1=a_1^*,\ldots,A_K=a_K^*} = Y$

   if $A_k = a_k^*$ for $k = 0, 1, \ldots, K$

2. Exchangeability: $Y^{A_0=a_0^*,A_1=a_1^*,\ldots,A_K=a_K^*} \perp\!\!\!\perp A_k \mid \bar{X}_k, \bar{A}_{k-1}$

   for $k = 0, 1, \ldots, K$

3. Positivity: $$\prod_{k=0}^{K} \frac{1\{A_k = a_k^*\}}{P(A_k = a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} < \infty$$

   for $k = 0, 1, \ldots, K$

Notation for histories of variables: $\bar{X}_k = (X_0, X_1, \ldots, X_k)$, $\bar{A}_k = (A_0, A_1, \ldots, A_k)$.

Time-dependent treatment interventions

Imposing a static regime, like 'always treat',

$$A_0 = 1, A_1 = 1, \dots, A_K = 1$$

may not always be realistic (or even feasible).

Time-dependent treatment interventions

Imposing a static regime, like 'always treat',

$$A_0 = 1, A_1 = 1, \ldots, A_K = 1$$

may not always be realistic (or even feasible).

Example (Robins 1986) Effects of exposure of chemicals on employees: Static regimes cannot be identified since subjects can only be exposed when at work.

Time-dependent treatment interventions

Imposing a static regime, like 'always treat',

$$A_0 = 1, A_1 = 1, \ldots, A_K = 1$$

may not always be realistic (or even feasible).

Example (Robins 1986) Effects of exposure of chemicals on employees: Static regimes cannot be identified since subjects can only be exposed when at work.

Another example Development of adverse effects or contraindications (e.g., pregnancy) can force a subject to stop an assigned treatment.

Time-dependent treatment interventions

But the positivity assumption dictates that the treatment level imposed by the intervention cannot in the observed data be deterministically assigned at any time point based on a subject's observed past.

3. Positivity:
$$\prod_{k=0}^{K} \frac{1\{A_k = a_k^*\}}{P(A_k = a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} < \infty$$
for $k = 0, 1, \ldots, K$

Time-dependent treatment interventions

What we can do $\Rightarrow$ change the question/intervention.

- 'Expose when at work'
- 'Treat until adverse event or contraindication happen'
- 'Initiate antidiabetic treatment when HbA1c level increases beyond some level'

Time-dependent treatment interventions

Dynamic treatment regimes

- A prespecified set of rules which assign treatment over time by responding to a patient's time-varying conditions.

Time-dependent treatment interventions

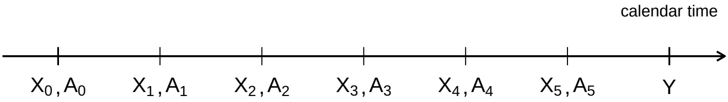
Dynamic treatment regimes

- A prespecified set of rules which assign treatment over time by responding to a patient's time-varying conditions.
- Mathematically, defined as function $\mathcal{S}_k(\bar{X}_k, \bar{A}_{k-1})$ that maps (a subset of) previous covariate/treatment values $\bar{X}_k, \bar{A}_{k-1}$ to a binary treatment assignment, e.g.,

$$\mathcal{S}_k(\bar{X}_k, \bar{A}_{k-1}) = \begin{cases} 1 & \text{if } X_k > \theta, \\ 0 & \text{if } X_k \leq \theta. \end{cases}$$

Time-dependent treatment interventions

2. Exchangeability: $Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*} \perp\!\!\!\perp A_k \mid \bar{X}_k, \bar{A}_{k-1}$ for $k = 0, 1, \ldots, K$

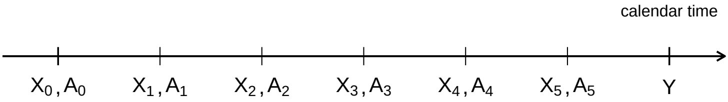


- X_k may be affected by earlier treatment decisions $A_{k-1}, \ldots, A_1, A_0$.
- X_k may be a confounder for the effect of $A_k, A_{k+1}, \ldots, A_K$ on Y.

Time-dependent treatment interventions

2. Exchangeability: $Y^{A_0=a_0^*, A_1=a_1^*, \dots, A_K=a_K^*} \perp\!\!\!\perp A_k \mid \bar{X}_k, \bar{A}_{k-1}$

   for $k = 0, 1, \dots, K$



- X_k may be affected by earlier treatment decisions $A_{k-1}, \dots, A_1, A_0$.
- X_k may be a confounder for the effect of $A_k, A_{k+1}, \dots, A_K$ on Y.

time-dependent confounding

Time-dependent treatment interventions

In presence of time-dependent confounding, "standard methods" may cause bias

- Multiple regression
- Random effects models
- Time-dependent Cox regression

The problem is that:

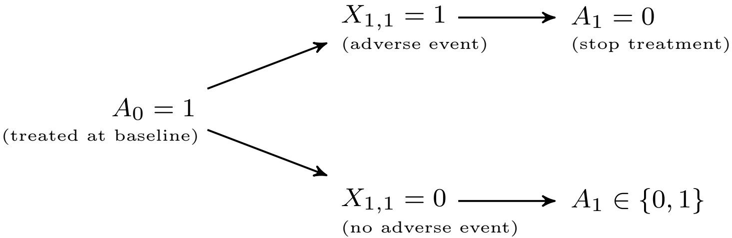
- If we control for X_k in our model, we will not capture the effect from earlier treatment decisions $A_{k-1}, \dots, A_1, A_0$ through X_k.
- But we have to control for X_k to assess the effect of $A_k, A_{k+1}, \dots, A_K$ on Y.

Time-dependent treatment interventions

A simulation setting

- $X_{0,1}, X_{0,2}, X_{0,3}$ are baseline covariates.
- $A_0 \in \{0,1\}$ is a randomized treatment indicator.
- $X_{1,1}, X_{1,2}$ are follow-up covariates.
- $A_1 \in \{0,1\}$ is a follow-up treatment decision.
- $Y \in \{0,1\}$ is the final outcome.

Time-dependent treatment interventions



- The variable $X_{1,1}$ is an indicator of an adverse event from the baseline treatment, an adverse event that causes treated subjects to switch from 'treatment' ($A_0 = 1$) to 'no treatment' ($A_1 = 0$).
- The variable $X_{1,2}$ is a marker of being likely to forget to take the medicin (or thinking it is too bothersome) which increases the probability of switching treatment as well.

Time-dependent treatment interventions

Say we are interested in the effects of different types of interventions:

1. The intention-to-treat (ITT) effect which only sets treatment at baseline and contrasts the two scenarios of being treated at baseline ($A_0 = 1$) and not being treated at baseline ($A_0 = 0$).
2. A static effect of being 'always treated' ($A_0 = A_1 = 1$) and 'never treated' ($A_0 = A_1 = 0$).
3. A dynamic effect of being treated at baseline ($A_0 = 1$) and only treated at follow-up if the adverse event has not happened, i.e., $X_{1,1} = 0$ — contrasted to being 'never treated' ($A_0 = A_1 = 0$).

Time-dependent treatment interventions

The true ITT average treatment effect:

```
ITT: -0.93%
```

The true static average treatment effect:

```
static: -6.33%
```

The true dynamic average treatment effect:

```
dynamic: -5.07%
```

Time-dependent treatment interventions

In the next lecture we will consider TMLE for estimation of all three types of effects.

For now, consider two 'naive approaches' to estimation of the static effect:

Time-dependent treatment interventions

In the next lecture we will consider TMLE for estimation of all three types of effects.

For now, consider two 'naive approaches' to estimation of the static effect:

1. Logistic regression of the outcome regressed on all treatment variables and covariates: Contrast means of the predictions under $A_0 = A_1 = 1$ to the mean of the predictions under $A_0 = A_1 = 0$.

Time-dependent treatment interventions

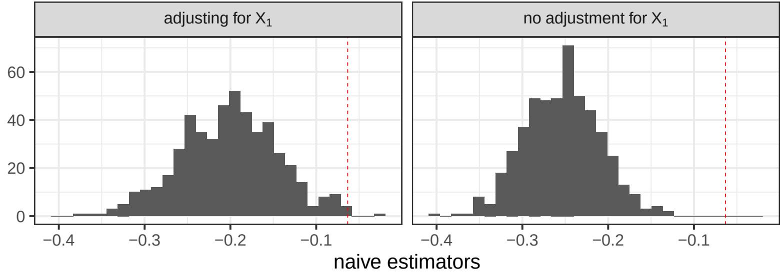
In the next lecture we will consider TMLE for estimation of all three types of effects.

For now, consider two 'naive approaches' to estimation of the static effect:

1. Logistic regression of the outcome regressed on all treatment variables and covariates: Contrast means of the predictions under $A_0 = A_1 = 1$ to the mean of the predictions under $A_0 = A_1 = 0$.
2. Logistic regression of the outcome regressed on baseline covariates and both treatment variables (leaving out follow-up covariates): Contrast means of the predictions under $A_0 = A_1 = 1$ to the mean of the predictions under $A_0 = A_1 = 0$.

Time-dependent treatment interventions

In a simulation study with $M = 500$ repetitions:



Both naive approaches give biased results — due to time-dependent confounding.

Practical:

Kreif et al. (2017) as an example

- Data structure, static and dynamic intervention, time-dependent treatment interventions.
- IP-weighting, g-formula, TMLE.

Questions for the paper that you should go over can be found in: **day3_practical1.pdf**.