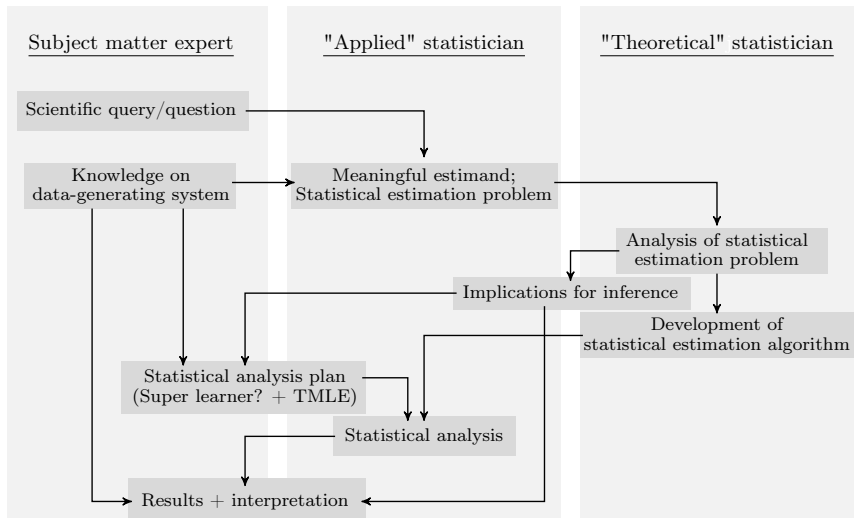


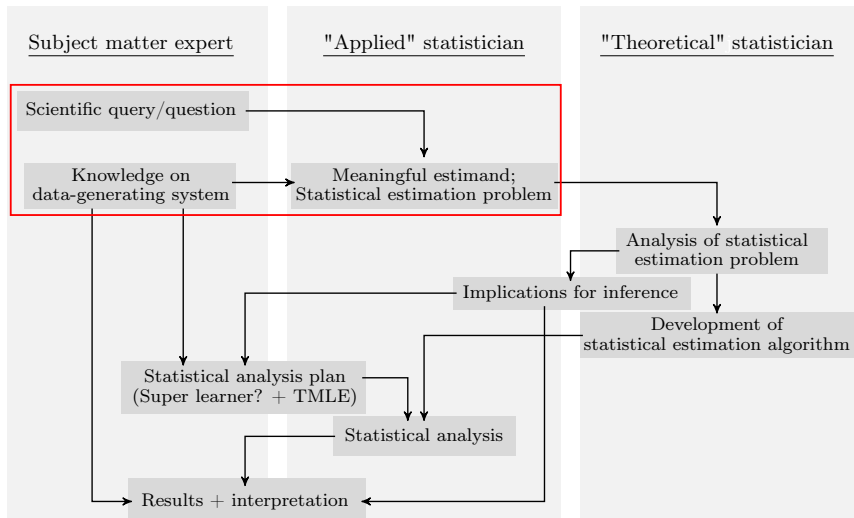
Day 1, Lecture 2

Properly defining the target parameter

Properly defining the target parameter



Properly defining the target parameter



Properly defining the target parameter

- ▶ A clearly defined goal as a starting-point for any analysis
 - ▶ necessary to talk about estimator performance
 - ▶ semiparametric/nonparametric efficiency theory (and TMLE) requires a clearly defined goal
- ▶ Brief introduction to the setting of a typical causal inference problem
 - ▶ example: average treatment effect
 - ▶ *model-free* and *estimator-free* definition of parameters

Moving targets with different logistic regression models

- ▶ $X \sim \text{Unif}(-2, 2)$
- ▶ $A \sim \text{Bernoulli}(0.5)$ (no confounding)
- ▶ $Y \in \{0, 1\}$

Say that the distribution of Y given X and A follows the parametric model:

$$\text{logit } \mathbb{E}[Y \mid A, X] = \beta_0 + \beta_A A + \beta_X^\top X^2$$

Moving targets with different logistic regression models

- ▶ $X \sim \text{Unif}(-2, 2)$
- ▶ $A \sim \text{Bernoulli}(0.5)$ (no confounding)
- ▶ $Y \in \{0, 1\}$

Say that the distribution of Y given X and A follows the parametric model:

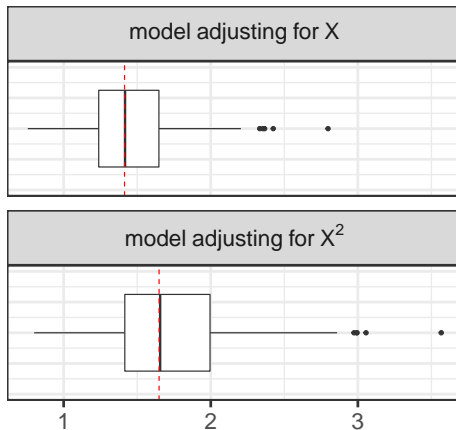
$$\text{logit } \mathbb{E}[Y \mid A, X] = \beta_0 + \beta_A A + \beta_X^\top X^2$$

The odds ratio $\exp(\beta_A)$ is a different parameter than $\exp(\alpha_A)$ in a different model:

$$\text{logit } \mathbb{E}[Y \mid A, X] = \alpha_0 + \alpha_A A + \alpha_X^\top X$$

Moving targets with different logistic regression models

- ▶ The variables X we include in the model to assess the effect of A on Y changes the parameter (conditional OR).
- ▶ Only one of the models can be true at a time.



The upper panel does not show a biased estimator, just an estimator targeting a different parameter (dashed red line).

Causal inference

What we obtain moving on to a causal inference setting: 1) An interpretable and relevant target of estimation, and 2) a model-free definition of a target parameter.

¹And, if you are already familiar, consider this a small repetition and introduction to the notation.

Causal inference

What we obtain moving on to a causal inference setting: 1) An interpretable and relevant target of estimation, and 2) a model-free definition of a target parameter.

- ▶ We are only going to go briefly over the "causal inference concepts",¹ but we need this part to very clear about with it is we are estimating.
 - ▶ For today and tomorrow we consider just the simple example where the target of estimation is the average treatment effect (ATE).
- ▶ For the causal inference notation, we follow the book by Hernán and Robins (which, if you are interested, you can find here: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf).
- ▶ I will leave out DAGs/SCMs.

¹And, if you are already familiar, consider this a small repetition and introduction to the notation.

Steps of the roadmap

- Step 1 Go from scientific question to target causal estimand (stated in the language of counterfactuals)
- Step 2 Assess whether we can go from target causal estimand to target statistical estimand = assess "identifiability"

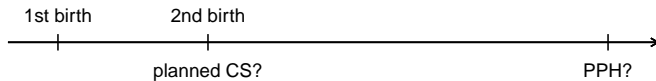
In a given data situation, we want to explicitly clarify:

1. Observed data
2. Causal model
3. Causal question and target causal estimand
4. Identifiability

An example we can have in the back of our minds

Scientific question:

Does having a planned cesarian section (intended cesarian section) among women who gave birth twice change the risk of postpartum haemorrhage (PPH) during the second delivery?



Goal: Translate this into a precise formulation of a statistical estimation problem.

Observed data

Observed data $O = (X, A, Y) \in \mathbb{R}^d \times \{0, 1\} \times \{0, 1\} = \mathcal{O}$

- * $X \in \mathbb{R}^d$ are covariates
ex: age at 2nd delivery, information of PPH at first delivery, ...
- * $A \in \{0, 1\}$ is a binary exposure variable (treatment decision)
ex: decision to have a planned cesarian section.
- * $Y \in \{0, 1\}$ is a binary outcome variable
ex: PPH (postpartum haemorrhage).

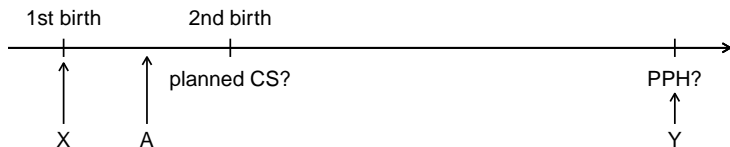
We observe a sample $O_1, \dots, O_n \stackrel{iid}{\sim} P_0 \in \mathcal{M}$, $n \in \mathbb{N}$.

\mathcal{M} is the set of all possible probability distributions for our data.

Observed data

Implicit assumptions for the data structure:²

- ▶ X are covariates known before the treatment decision A was made
- ▶ Outcome Y is observed after treatment decision was made



²This ordering could also be encoded in a structural causal model.

Observed data

Our statistical model \mathcal{M} for P_0 contains possible distributions P for the observed data O .

The density p of $P \in \mathcal{M}$ can be factorized into:

$$p(o) = \mu_Y(y \mid a, x) \pi(a \mid x) \mu_X(x),$$

- ▶ $\mu_Y(y \mid A, X) = P(Y = y \mid A, X)$
- ▶ $\pi(a \mid X) = P(A = a \mid X)$
- ▶ μ_X is the marginal density of X (with respect to an appropriate dominating measure)

Observed data

Our statistical model \mathcal{M} for P_0 contains possible distributions P for the observed data O .

The density p of $P \in \mathcal{M}$ can be factorized into:

$$p(o) = \mu_Y(y \mid a, x) \pi(a \mid x) \mu_X(x),$$

- ▶ $\mu_Y(y \mid A, X) = P(Y = y \mid A, X)$
- ▶ $\pi(a \mid X) = P(A = a \mid X)$
- ▶ μ_X is the marginal density of X (with respect to an appropriate dominating measure)

We assume that \mathcal{M} is a nonparametric model.

- * Throughout, we make *no parametric restrictions* on μ_Y, μ_X .
- * We could impose some parametric structure on π , but let us assume that we do not.

Operators on functions of the observed data³

For a function $h : \mathcal{O} \rightarrow \mathbb{R}$ and distribution P

$$Ph = \mathbb{E}_P[h(O)] = \int h dP = \int_{\mathcal{O}} h(o) dP(o)$$

where $\mathcal{O} = \mathbb{R}^d \times \{0, 1\} \times \{0, 1\}$ is the sample space of $O = (X, A, Y)$.

³van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press.

Confounding

How can we define a causal effect?

The contrast $\mathbb{E}_P[Y \mid A = 1] - \mathbb{E}_P[Y \mid A = 0]$ tells us about the risk difference in the two exposure groups.

Any such difference is likely due to other factors than the decision to initiate treatment or not

- * the exposure decision is **confounded**.

Counterfactuals

To answer a causal question, we ideally want to know

Scenario 1 What happened to a subject had they been exposed?

Scenario 2 What would have happened to the same subject had they not been exposed?

We imagine a model with two outcomes for each subject:

- ▷ a variable Y^1 corresponding to scenario 1, and
- ▷ a variable Y^0 corresponding to scenario 2

= the "counterfactuals" (aka **potential outcomes**).

Counterfactuals

- * Y^1 = outcome if exposed
- * Y^0 = outcome if not exposed

We use the counterfactual outcomes to define precisely what a causal effect is:

- on the individual level, $Y^1 = 1$ and $Y^0 = 0$ for a particular subject would tell us that this subject would experience outcome under exposure and not otherwise
- on the population level, $\mathbb{E}_P[Y^1] \neq \mathbb{E}_P[Y^0]$ tells us that the risk changes depending on whether exposed or not

Target causal estimand: Average causal effect (ATE)

The average causal effect⁴ (ATE/ACE) measures the average effect in the population

$$\text{ATE} = \mathbb{E}_P[Y^1] - \mathbb{E}_P[Y^0]$$

- It is interpreted as the difference in risk had everyone in the population been exposed and had everyone in the population been unexposed.

⁴or average treatment effect.

Identifiability (estimating the causal effect from observational data)

Can we estimate the causal effect from the observed data?

- ▷ only Y^1 or Y^0 is observed for each individual.

Identifying $\mathbb{E}_P[Y^1] - \mathbb{E}_P[Y^0]$

= write $\mathbb{E}_P[Y^1] - \mathbb{E}_P[Y^0]$ as a parameter of the observed data distribution.

requires three overall assumptions (identifiability assumptions).

Identifiability (estimating the causal effect from observational data)

1. Consistency: $Y^a = Y$ if $A = a$, $a = 0, 1$

- ▶ Requires that the "treatment intervention" is well-defined and no interference between subjects.
- ▶ Example of a violation: effect of vaccines (one subject's effect of a vaccine depends on whether other subjects are vaccinated or not).

Identifiability (estimating the causal effect from observational data)

1. Consistency: $Y^a = Y$ if $A = a$, $a = 0, 1$

- ▶ Requires that the "treatment intervention" is well-defined and no interference between subjects.
- ▶ Example of a violation: effect of vaccines (one subject's effect of a vaccine depends on whether other subjects are vaccinated or not).

2. Exchangeability: $Y^a \perp\!\!\!\perp A \mid X$, for $a = 0, 1$

- ▶ Conditional on covariates, the exposed group tells us what would happen to the unexposed if they had been exposed and vice versa.
- ▶ Requires that there is **no unmeasured confounding**.

Identifiability (estimating the causal effect from observational data)

1. Consistency: $Y^a = Y$ if $A = a$, $a = 0, 1$

- ▶ Requires that the "treatment intervention" is well-defined and no interference between subjects.
- ▶ Example of a violation: effect of vaccines (one subject's effect of a vaccine depends on whether other subjects are vaccinated or not).

2. Exchangeability: $Y^a \perp\!\!\!\perp A \mid X$, for $a = 0, 1$

- ▶ Conditional on covariates, the exposed group tells us what would happen to the unexposed if they had been exposed and vice versa.
- ▶ Requires that there is **no unmeasured confounding**.

3. Positivity: $P(A = a \mid X) > 0$ for $a = 0, 1$ and almost surely all X

- ▶ We cannot investigate the effect of an intervention that was never "tested" in the observed data (conditional on covariates X).

Identifiability (estimating the causal effect from observational data)

Under these assumptions:

$$\begin{aligned}\mathbb{E}_P[Y^1] - \mathbb{E}_P[Y^0] \\&= \mathbb{E}_P[\mathbb{E}_P[Y^1 | X] - \mathbb{E}_P[Y^0 | X]] \\&\stackrel{2.}{=} \mathbb{E}_P[\mathbb{E}_P[Y^1 | A = 1, X] - \mathbb{E}_P[Y^0 | A = 0, X]] \\&\stackrel{1.}{=} \mathbb{E}_P[\mathbb{E}_P[Y | A = 1, X] - \mathbb{E}_P[Y | A = 0, X]] \\&= \Psi(P)\end{aligned}$$

(3. (positivity) ensures that the conditional expectations are well-defined).

Goal achieved: Right hand side is expressed only in terms of observable quantities.

Identifiability (estimating the causal effect from observational data)

Under the assumptions:

$$\mathbb{E}_P[Y^1] - \mathbb{E}_P[Y^0] = \underbrace{\mathbb{E}_P[\mathbb{E}_P[Y \mid A = 1, X] - \mathbb{E}_P[Y \mid A = 0, X]]}_{(*)} = \Psi(P),$$

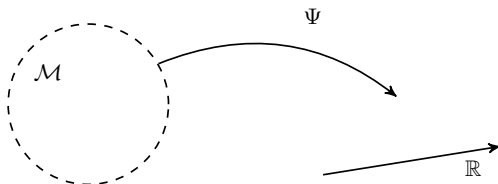
for any $P \in \mathcal{M}$.

In our statistical analysis, we proceed with (*).

"Causal inference part" is over.

Target statistical estimand

Now we are exactly in the situation we wanted:



Average treatment effect (ATE)

- ▶ $O = (X, A, Y) \in \mathbb{R}^d \times \{0, 1\} \times \{0, 1\}$
- ▶ The ATE is defined for $P \in \mathcal{M}$ as

$$\Psi(P) = \mathbb{E}_P[\mathbb{E}_P[Y \mid A = 1, X] - \mathbb{E}_P[Y \mid A = 0, X]]$$

Target statistical estimand: g-formula

We can rewrite the target parameter as:

$$\begin{aligned}\Psi(P) &= \mathbb{E}_P[\mathbb{E}_P[Y \mid A = 1, X] - \mathbb{E}_P[Y \mid A = 0, X]] \\ &= \mathbb{E}_P[f(1, X) - f(0, X)] \\ &= \int_{\mathbb{R}^d} (f(1, x) - f(0, x)) d\mu_X(x) = \tilde{\Psi}(f, \mu_X) \quad (*)\end{aligned}$$

where

$$f(a, x) = \mathbb{E}[Y \mid A = a, X = x]$$

and μ_X is the marginal distribution of X .

We refer to this as the **g-formula**.

Target statistical estimand: IP-weighting

We can also rewrite the target parameter as:

$$\begin{aligned}\Psi(P) &= \int_{\mathbb{R}^d} (f(1, x) - f(0, x)) d\mu_X(x) & (*) \\ &= \int_{\mathbb{R}^d} \sum_{y=0,1} y (\mu_Y(y \mid 1, x) - \mu_Y(y \mid 0, x)) d\mu_X(x) \\ &= \int_{\mathbb{R}^d} \sum_{y=0,1} \sum_{a=0,1} y (a\mu_Y(y \mid a, x) - (1-a)\mu_Y(y \mid a, x)) d\mu_X(x) \\ &= \int_{\mathbb{R}^d} \sum_{y=0,1} \sum_{a=0,1} \left(\frac{ay}{\pi(a \mid x)} - \frac{(1-a)y}{\pi(a \mid x)} \right) \mu_Y(y \mid a, x) \pi(a \mid x) d\mu_X(x) \\ &= \tilde{\Psi}_{\text{ipw}}(\pi, p) \quad (**)\end{aligned}$$

where $\pi(a \mid x) = P(A = 1 \mid X = x)$.

Target statistical estimand

The **g-formula**:

$$\begin{aligned}\tilde{\Psi}(f, \mu_X) &= \int_{\mathbb{R}^d} (f(1, x) - f(0, x)) d\mu_X(x) \\ &= \mathbb{E}_P[f(1, X) - f(0, X)].\end{aligned}\quad (*)$$

The **IP-weighted formula**:

$$\begin{aligned}\tilde{\Psi}_{\text{ipw}}(\pi, p) &= \int_{\mathbb{R}^d} \sum_{a=0,1} \sum_{y=0,1} \left(\frac{ay}{\pi(a | x)} - \frac{(1-a)y}{\pi(a | x)} \right) dP(o) \\ &= \mathbb{E}_P \left[\frac{AY}{\pi(1 | X)} - \frac{(1-A)Y}{\pi(0 | X)} \right]\end{aligned}\quad (**)$$

- ▶ f and (the average over) μ_X are **nuisance parameters** for the g-formula.
- ▶ π and (the average over) p are **nuisance parameters** for the IP-weighted formula.

Target statistical estimand

Yet another representation of the target parameter is

$$\begin{aligned}\tilde{\Psi}_{\text{ee}}(f, \pi, p) &= \int_{\mathbb{R}^d} \sum_{a=0,1} \sum_{y=0,1} \left\{ \left(\frac{a}{\pi(a|x)} - \frac{1-a}{\pi(a|x)} \right) (y - f(a, x)) \right. \\ &\quad \left. + f(1, x) - f(0, x) \right\} p_Y(y | a, x) \pi(a | x) d\mu_X(x) \\ &= \mathbb{E}_P \left[\left(\frac{A}{\pi(A|X)} - \frac{1-A}{\pi(A|X)} \right) (Y - f(A, X)) + f(1, X) - f(0, X) \right]\end{aligned}$$

- f , π and (the average over) p are **nuisance parameters** for this parametrization.

Target statistical estimand

Yet another representation of the target parameter is

$$\begin{aligned}\tilde{\Psi}_{\text{ee}}(f, \pi, p) &= \int_{\mathbb{R}^d} \sum_{a=0,1} \sum_{y=0,1} \left\{ \left(\frac{a}{\pi(a|x)} - \frac{1-a}{\pi(a|x)} \right) (y - f(a, x)) \right. \\ &\quad \left. + f(1, x) - f(0, x) \right\} p_Y(y | a, x) \pi(a | x) d\mu_X(x) \\ &= \mathbb{E}_P \left[\left(\frac{A}{\pi(A|X)} - \frac{1-A}{\pi(A|X)} \right) (Y - f(A, X)) + f(1, X) - f(0, X) \right]\end{aligned}$$

- f , π and (the average over) p are **nuisance parameters** for this parametrization.

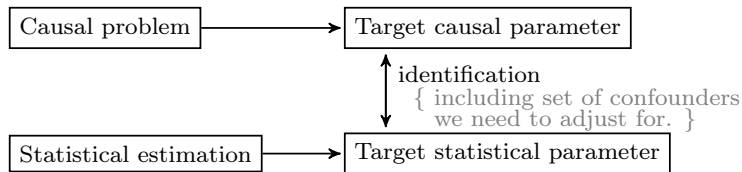
SMALL EXERCISE:

1. Note that $\tilde{\Psi}_{\text{ee}}(f, \pi, p) = \tilde{\Psi}(f, \mu_X) + [\text{an extra term}]_1$. Show that it can also be written as $\Psi_{\text{ee}}(f, \pi, p) = \tilde{\Psi}_{\text{ipw}}(\pi, p) + [\text{an extra term}]_2$.
2. Show that $\tilde{\Psi}_{\text{ee}}(f, \pi, p) = \mathbb{E}_P[Y^1] - \mathbb{E}_P[Y^0]$ under the identifiability assumptions (consistency, exchangeability and positivity).

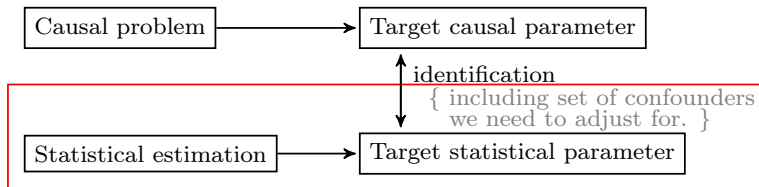
Summary: Properly defining the target

- ▶ Causal parameter (now fixed).
- ▶ Causal model: How do the observed variables affect one another?
 - ▶ are the covariates we observe sufficient to remove confounding? which variables **do we need to adjust for** to make treatment groups comparable?
- ▶ Identifiability
 - ▶ identifiability assumptions allow us to write causal parameter as statistical parameter.
 - ▶ the assumptions may not hold, but we can state and discuss them.
- ▶ Statistical parameter
 - ▶ Statistical interpretation: The average effect in the population, standardized to the distribution of covariates.

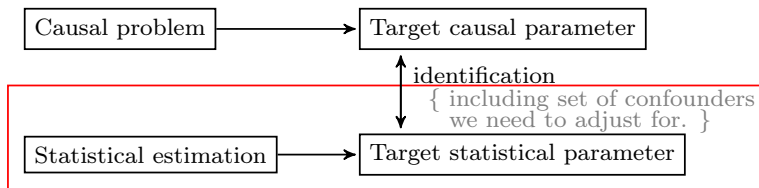
Summary: Properly defining the target



Summary: Properly defining the target



Summary: Properly defining the target



- ▶ one estimator is not more causal than another.
- ▶ different estimators are based on different nuisance parameters and have different statistical properties (bias/variance).

On a sidenote: Other simple causal parameters

We focus on the ATE as an example of a causal parameter.

But note that other simple causal parameters can be constructed from $\mathbb{E}[Y^1]$ and $\mathbb{E}[Y^0]$.

Like:

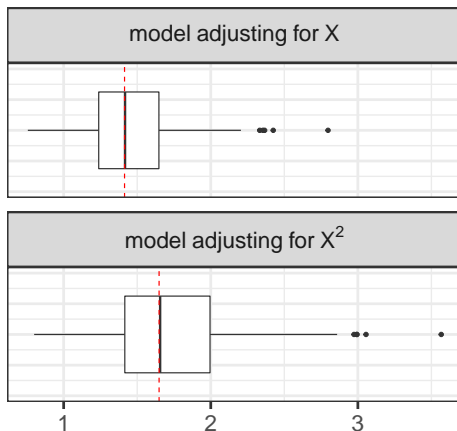
$$\psi^{\text{RR}}(P) = \frac{\mathbb{E}[Y^1]}{\mathbb{E}[Y^0]},$$

or,

$$\psi^{\text{OR}}(P) = \frac{\mathbb{E}[Y^1]/(1 - \mathbb{E}[Y^1])}{\mathbb{E}[Y^0]/(1 - \mathbb{E}[Y^0])}.$$

On a sidenote: Other simple causal parameters

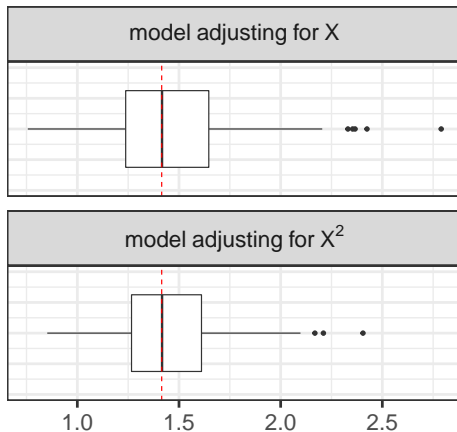
$\log(\text{OR})$ as a **regression coefficient** is a moving target in different logistic regression models:



The upper panel does not show a biased estimator, just an estimator targeting a different parameter (dashed red line).

On a sidenote: Other simple causal parameters

The corresponding **causal odds ratio** is a fixed target — and the target does not change depending on adjustment for X or X^2 :



... but these are different statistical estimators, and they have different statistical properties.

Last slide of this lecture

Summarizing this lecture:

- ▶ take 5 minutes to write down 3–10 keywords/concepts/formulas from this lecture;
- ▶ discuss the keywords with the person sitting next to you, and explain their significance in the overall targeted learning framework (5 min).