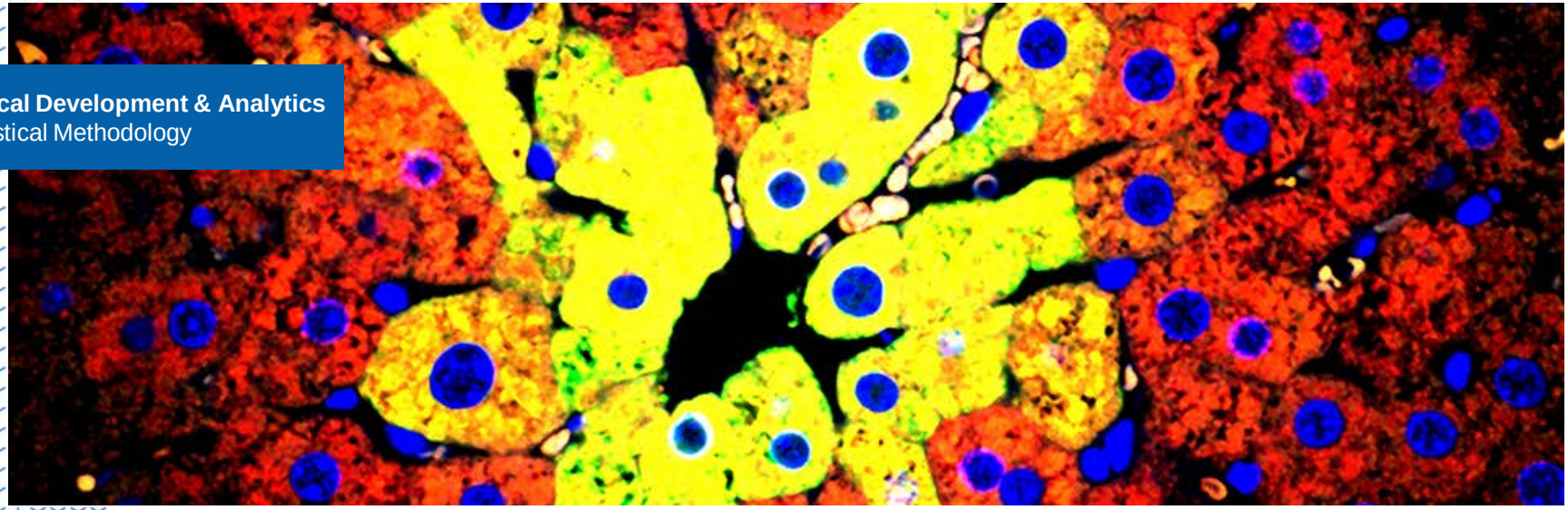
Clinical Development & Analytics
Statistical Methodology



A Gentle Introduction to Causal Inference in View of the ICH E9 Addendum on Estimands

Björn Bornkamp, Heinz Schmidli, Dong Xi

ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop
September 22, 2020

NOVARTIS | Reimagining Medicine

Disclaimer

The views and opinions expressed in this presentation and on the slides are solely those of the presenter and not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein.

Acknowledgements

- Novartis
  - Simon Newsome, Baldur Magnusson, Angelika Caputo
- Harvard University
  - Alex Ocampo
- Johns Hopkins University
  - Daniel Scharfstein

Agenda

10:00 – 11:40 AM

- Introduction to causal effects & potential outcomes (Heinz Schmidli)
- Relation to questions and concepts encountered in randomized clinical trials (Björn Bornkamp)

12:00 – 1:30 PM

- Standardization & inverse probability weighting (Dong Xi)

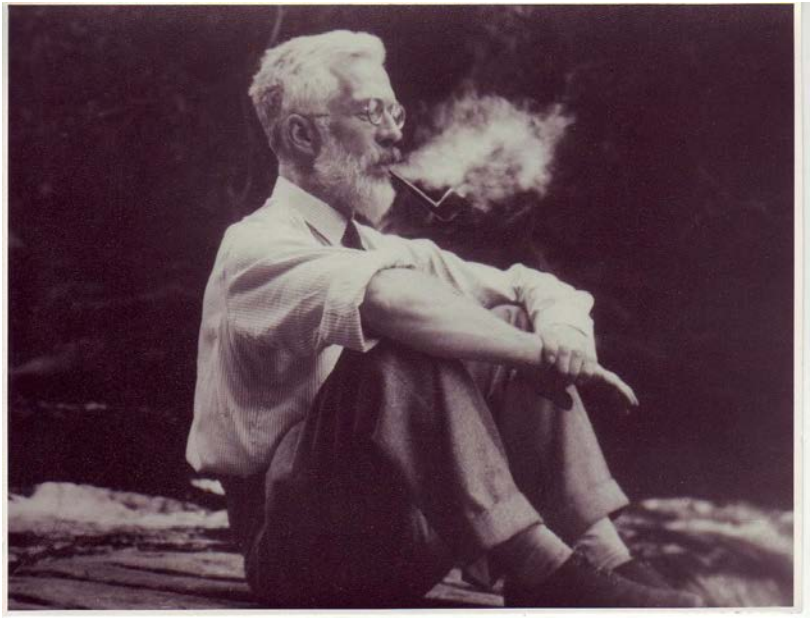
Part 1:
Introduction to causal effects and potential outcomes

Outline

- Causal effects
- Potential outcomes
- Causal estimands
- Causal inference
- Clinical development
- Conclusions

Causal effects

Does smoking cause lung cancer?



Cancer and Smoking

The curious associations with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer.

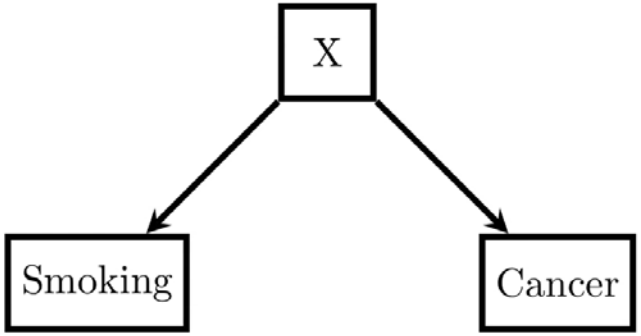
Ronald A. Fisher
Nature 1958;182(4635):596.

Causal effects

Directed Acyclic Graph (DAG) to express causal relationships

Smoking causes cancer

Patient characteristic X causes both smoking and cancer



Poll question 1

Do you believe that smoking causes lung cancer?

- YES
- NO

Why?

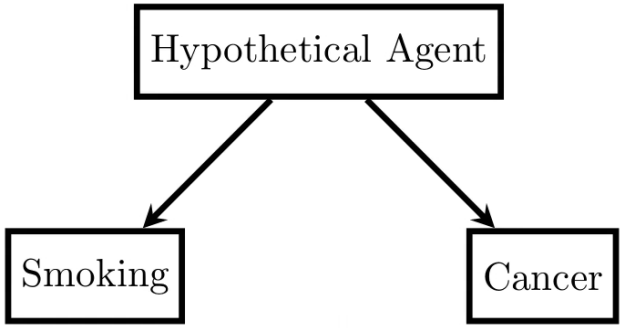
Causal effects

Smoking and lung cancer: recent evidence and a discussion of some questions*

Jerome Cornfield,[1] William Haenszel,[2] E. Cuyler Hammond,[3] Abraham M. Lilienfeld,[4] Michael B. Shimkin[5] and Ernst L. Wynder[6]

J. Nat. Cancer Inst. **22:**173–203, 1959

"The magnitude of the excess lung-cancer risk among cigarette smokers is so great that the results can not be interpreted as arising from an indirect association of cigarette smoking with some other agent or characteristic, since this hypothetical agent would have to be at least as strongly associated with lung cancer as cigarette use; no such agent has been found or suggested."



Causal effects

Smoking and lung cancer: recent evidence and a discussion of some questions*

Jerome Cornfield,[1] William Haenszel,[2] E. Cuyler Hammond,[3] Abraham M. Lilienfeld,[4] Michael B. Shimkin[5] and Ernst L. Wynder[6]

J. Nat. Cancer Inst. **22:**173–203, 1959

"The consistency of all the epidemiologic and experimental evidence also supports the conclusion of a causal relationship with cigarette smoking, while there are serious inconsistencies in reconciling the evidence with other hypotheses which have been advanced."

Causal effects

Association, Prediction, Causality

- Carrying a lighter is strongly associated with lung cancer
- Whether or not somebody carries a lighter is predictive of lung cancer
- But this is not a causal relationship!

In some settings, having a good predictive model may be sufficient.

In others, causality is of main interest

Potential outcomes

Mathematical language needed to express causal questions quantitatively, and to make causal inference

Potential outcomes framework provides this language

- Neyman (1923), Rubin (1974)
- Widely accepted (Robins, Pearl, Hernán, ...), with few exceptions (e.g. Dawid)

Requires a thought experiment:

What would outcomes be if action 1 vs 2 was taken?

(Some authors use the term counterfactuals rather than potential outcomes. Others use the term counterfactual for the potential outcome not observed)

Potential outcomes

A clinical study of Test (Z=1) vs Control (Z=0) treatment

- **Population**: patients with small-cell lung cancer
- **Variable**: time-to-death Y in years, from time of treatment assignment

Potential outcome framework

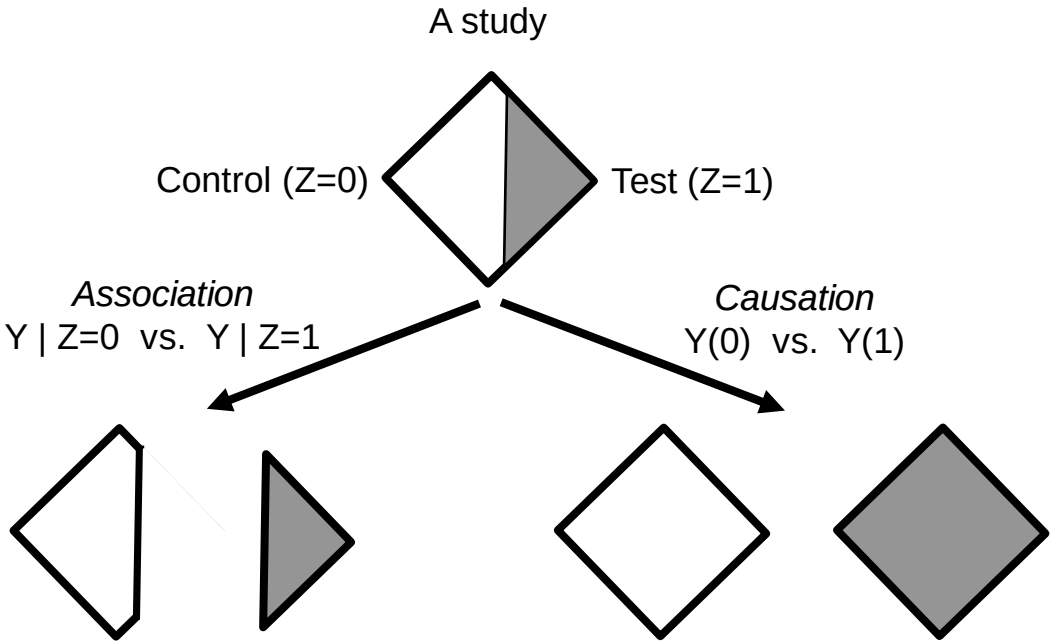
#	Patient	Y(1)	Y(0)
1	Adam	2	1
2	Bruce	5	7
3	...		

For the whole population

Potential outcomes Y(1), Y(0):

- Y(1): how long the patient would live if assigned to Test (Z=1)
- Y(0): how long the patient would live if assigned to Control (Z=0)

Potential outcomes



Hernán, Robins (2020)

The critical requirement is that to be a causal effect, the comparison must be a comparison of $Y_i(1)$ and $Y_i(0)$ for a common set of units. More formally, a causal effect must be a comparison of the ordered sets $\{Y_i(1), i \in S\}$ and $\{Y_i(0), i \in S\}$, not $\{Y_i(1), i \in S_1\}$ and $\{Y_i(0), i \in S_0\}$, where S_1 and S_0 are not equal.

Rubin (2006) JASA

Causal estimands

Potential outcome framework

#	Patient	Y(1)	Y(0)
1	Adam	2	1
2	Bruce	5	7
	...		

Y(1), Y(0): how long patient would live if assigned to Test (1) or Control (0)

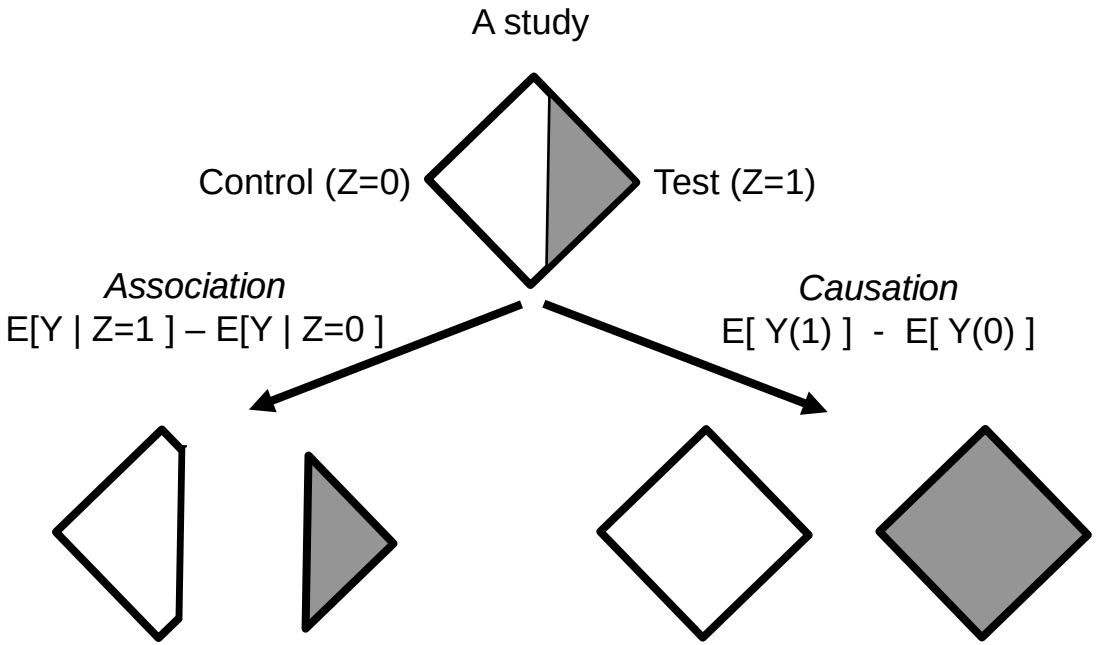
Treatment effect measure: population causal estimand

$$\text{e.g.} \quad E[\,Y(1) - Y(0)\,] = E[\,Y(1)\,] - E[\,Y(0)\,]$$

Many alternative causal estimands, e.g.

- $E[\log\{\,Y(1)\,\}] - E[\log\{\,Y(0)\,\}] = E[\log\{\,Y(1)/Y(0)\,\}]$ Accelerated Life Time
- $\text{Median}\{\,Y(1)\,\} \;/\; \text{Median}\{\,Y(0)\,\}$ Median survival times

Causal estimands



Causal estimands

US National Academy of Science (2010)

“The trial protocol should explicitly define

- the objective(s) of the trial;
- the associated primary outcome or outcomes;
- how, when, and on whom the outcome or outcomes will be measured;
- The measures of intervention effects, that is, the **causal estimands** of primary interest.

These measures should be meaningful for all study participants, and estimable with minimal assumptions.”

Causal estimands

ICH E9(R1) Estimands and Sensitivity Analysis in Clinical Trials (2019)

Aligned with causal reasoning, although term “causal” not used.

“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).”*

Causal inference

For each patient, at most one of Y(1) or Y(0) observed.

'Fundamental problem of causal inference' Holland (1986)

		Potential outcomes		Observed			
#	Patient	Y(1)	Y(0)	Treatment Z	Y	Y(1)	Y(0)
1	Adam	2	1	1	2	2	?
2	Bruce	5	7	0	7	?	7
	...						

Population causal estimand, e.g. E[Y(1)] – E[Y(0)]

In randomized controlled trials (RCTs):

$$E[Y(1)] – E[Y(0)] = E[Y \mid Z=1] – E[Y \mid Z=0]$$

Causal inference

Generally: $E[Y(1)] – E[Y(0)] \neq E[Y \mid Z=1] – E[Y \mid Z=0]$

Models/assumptions needed for statistical inference on the causal estimand (causal inference):

- Model for assignment of treatment to patients
- Model for potential outcomes

Essential for observational studies, but also for some scientific questions in RCT's due to post-baseline (intercurrent) events (examples following)

Two simple approaches for estimation of causal estimands discussed later: standardization and inverse probability weighting

Q & A

Clinical development

- Causal questions are central to clinical development
- Randomization facilitates causal inference
- Complex questions regarding causality may arise in RCTs
- It is important to recognize these
- We will discuss some examples in the following

ITT analysis
Poll question 2

Double-blind randomized trial with continuous endpoint Y at week 12
→ Patients randomized to daily doses of the investigational treatment (Z=1) or control (Z=0).

At the end of the trial one calculates

$$\text{Mean}[Y_i \mid Z_i = 1] - \text{Mean}[Y_i \mid Z_i = 0],$$

i.e., the difference in means between patients randomized to Z=1 and Z=0, regardless of how frequently the patient takes the treatment.

Does this quantity estimate a causal effect?

- YES
- NO

ITT analysis

Definition of causal effect:
Comparison of $\{Y_i(1), i \in S\}$ and $\{Y_i(0), i \in S\}$, and not
$\{Y_i(1), i \in S_1\}$ with $\{Y_i(0), i \in S_0\}$, where S_1 and S_0 are not equal.

- Recall definition of causal effect
  - Z randomized → patients with Z=1 and Z=0 constitute „the same“ population
    - (formally $Y(0), Y(1) \perp Z$ see next session)
  - Yes, this is estimating a causal effect
  - Estimand of the analysis is: E[Y(1)] – E[Y(0)]
    - Can be estimated by the difference in observed means

ITT analysis

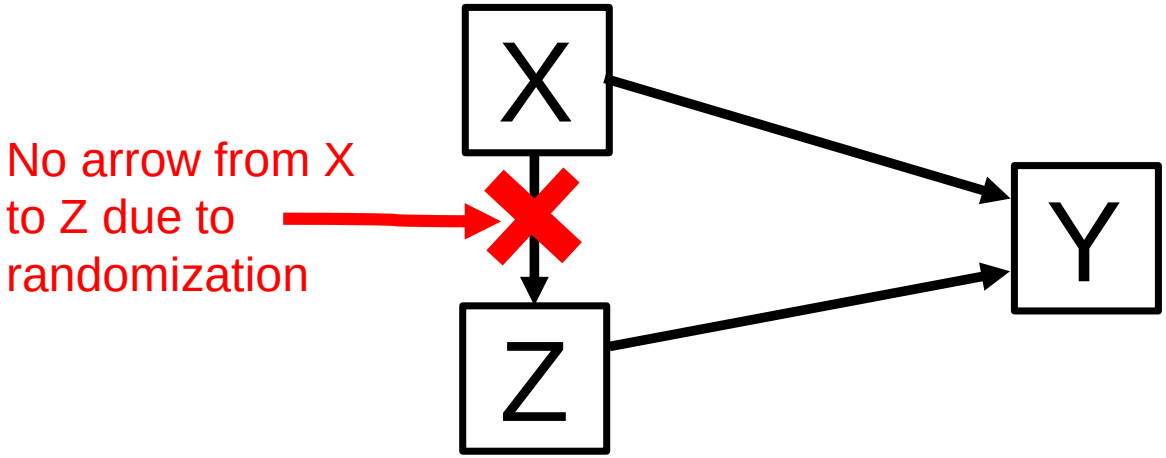
But: The causal effect of what?

- Causal effect of being randomized to a treatment
  → Does this correspond to a clinically relevant question?
  Depends...
  - on whether post-baseline events & subsequent actions will also occur in the same way in a real-life setting, on level of adherence to treatment, ...
- ITT does not estimate the effect of treatment: „had everyone adhered“
  - Different question!
- Causal inference requires clear definition of what „treatment“ constitutes (SUTVA, consistency assumptions, see slide notes)
  - If there are multiple versions of „treatment“ potential outcomes not well-defined

ITT analysis

- Final ICH E9 addendum: „Treatment“ is an additional estimand attribute.
  - For treatment policy (ITT) strategy, intercurrent events become part of „treatment“ attribute
  - Will (hopefully) lead to more transparency
- No longer
  - Treatment: 150mg twice daily
- Now
  - Treatment: Initiate 150mg twice daily + optional rescue medication + optional switch to another treatment if an adverse event requiring treatment discontinuation occurs.
- Clinical relevance of treatment policy strategy for dealing with intercurrent events (rescue medication, AE above) needs to be assessed on a case-by-case basis

Randomization in DAGs



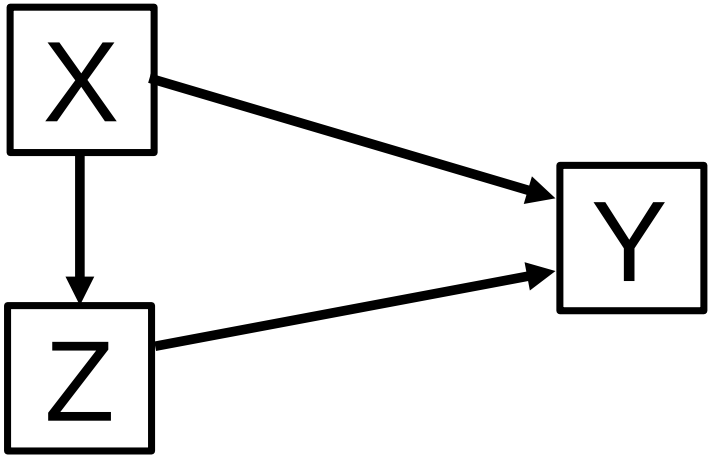
Z – treatment
Y – outcome
X – patient covariates

As there is no arrow pointing into Z in this DAG, all the association between Z and Y, must be due to the causal effect of Z on Y (i.e. association = causation in this DAG)

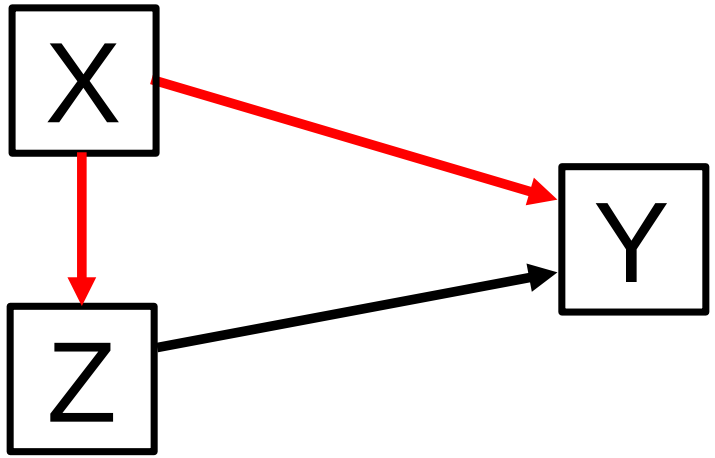
Observational data

Now assume the treating physician assigns treatment (Z) based on baseline severity of the disease X.
Variable: time-to-death Y in years since start of treatment
Estimand: E[Y(1)] – E[Y(0)]

DAG:

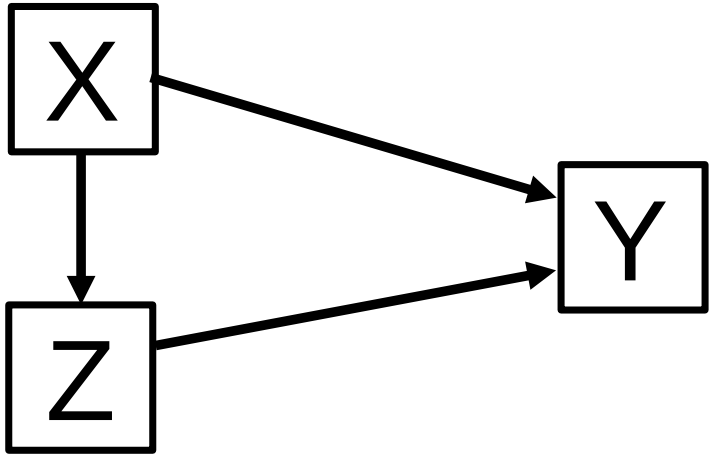


Observational data



- Problem in this DAG
  Non-causal path between Y and Z (with X pointing into Z): X is a confounder and will induce (non-causal) association between Z and Y

Observational data example



- In this example estimand *cannot* be estimated by a comparison of the observed means: Mean[Y_i | Z_i = 1] – Mean[Y_i | Z_i = 0]
- Populations on the two treatment arms are different: Observed difference can be due to difference in treatment or difference in population

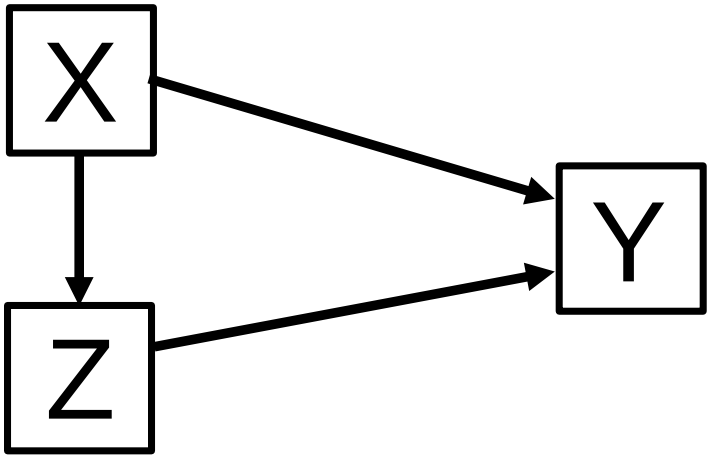
Observational data

Now assume the treating physician assigns treatment (Z) based on baseline severity of the disease X.
Variable: time-to-death Y in years since start of treatment

#	Patient	Y(1)	Y(0)	X	Z	Y
1	Adam	3	2	High	1	3
2	Bruce	4	4	Low	0	4
3	Carl	2	1	High	1	2
4	Dave	3	4	Low	0	4
...						
Population Mean		3.00	2.75			

$\text{Mean}[Y_i \mid Z_i=1] = (3+2)/2 = 2.50$
$\text{Mean}[Y_i \mid Z_i=0] = (4+4)/2 = 4.00$

Observational data



- Unequal population on the treatment arms
  - Cannot establish the causal effect according to the definition of a causal effect mentioned earlier
- But DAG implies: Only X influences outcome Y
  - To achieve „the same population“ across treatment arms, it is enough to balance X across treatment groups (e.g. with standardization or inverse probability weighting)
- No unmeasured confounders assumption
  - More formally: Conditional independence assumption $Y(0), Y(1) \perp Z \mid X$ (also called conditional ignorability or conditional exchangeability). „Within levels of X randomized assignment of treatment Z“

Analyses based on per-protocol set Poll question 3

Now again assume the setting of a randomized clinical trial.

Let A=1 and A=0 be inclusion or exclusion in the per-protocol set. Assume we calculate

$$\text{Mean}[Y_i \mid Z_i = 1, A_i = 1] - \text{Mean}[Y_i \mid Z_i = 0, A_i = 1],$$

i.e., the difference in means between patients that adhered to the protocol.

Does this quantity estimate a causal effect?

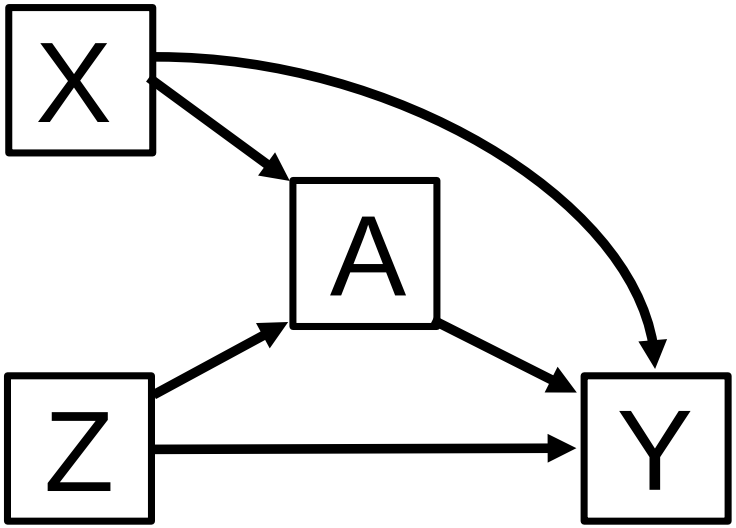
- YES
- NO

Analyses based on per-protocol set

- Let A(1) and A(0) be the potential outcomes of protocol adherence
- In potential outcomes we would compare
  - $\text{Mean}[Y(1)_i \mid Z_i = 1, A(1)_i = 1] - \text{Mean}[Y(0)_i \mid Z_i = 0, A(0)_i = 1]$
- Population with A(1) = 1 (protocol adherers under treatment) and A(0) = 1 (protocol adherers under placebo) can be different
  - **Not a causal effect!**
  - Per-protocol analyses are discouraged in the final ICH E9 addendum.

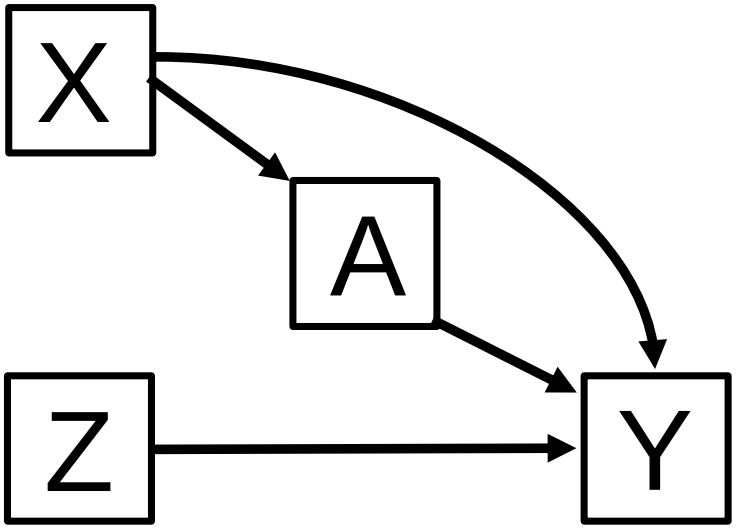
Analyses based on per-protocol set

Z – treatment
Y – outcome
A – protocol adherence
X – patient covariates



Analyses based on per-protocol set

Z – treatment
Y – outcome
A – protocol adherence
X – patient covariates



If this DAG would be true, the per-protocol analysis would target a causal effect as A is unaffected by Z, so that A(1) = A(0)

Analyses based on per-protocol set

What would be an estimand mimicking the idea of per-protocol analyses?

- Difference in means in patients that would adhere to the protocol under (i) control **and** investigational treatment or (ii) only the investigational treatment

(i) $E[Y(1) \mid Z = 1, A(1) = 1, A(0) = 1] - E[Y(0) \mid Z = 0, A(1) = 1, A(0) = 1]$

(ii) $E[Y(1) \mid Z = 1, A(1) = 1] - E[Y(0) \mid Z = 0, A(1) = 1]$

- Principal stratum strategy

Analyses based on per-protocol set

- Estimand (i)
  - only either A(0) or A(1) observed for every patient (never both)
  - see Lou et al. (2019) for an interesting approach for testing & estimation in the context of bioequivalence trials
- Estimand (ii)
  - A(1) not observed on control arm
  - can be harder to justify: Is Y(0) defined if A(0) = 0?

Alternative estimand based on hypothetical strategy

- Difference in means “had all patients adhered”
  - Can be clinically hard to justify (depending on reason for protocol non-adherence)
  - Final ICH E9 addendum discourages the hypothetical strategy for scenarios that would change the patients’ behaviors (rather than change the study design)

CANTOS trial

- Canakinumab is a monoclonal anti-body blocking interleukin-1β resulting in decreased inflammation
- Reducing inflammation without affecting lipid levels may reduce the risk of cardiovascular disease
- Level of inflammation measured by hsCRP
  - Inflammatory marker: Known to have a prognostic effect: Larger values are associated with a higher risk of CV related events

CANTOS trial

- CANTOS trial (Ridker et al. (2017))
  - Tested whether canakinumab leads to reduction in the risk for CV events.
  - Population: With previous myocardial infarction and hsCRP > 2 mg/L
  - Treatment groups: 3 dose groups of canakinumab versus placebo
  - Primary outcome: Time to first major adverse cardiovascular events (MACE)
- Result (for 150mg dose, focus only on this in what follows)
  - Hazard ratio of 0.85 (significant in multiple test strategy)
- Idea
  - Patients for whom the hsCRP is not lowered after 3 months might have a reduced benefit from canakinumab (and those where a lowering is seen, an increased benefit)

CANTOS trial

An analysis comparing patients with hsCRP < 2 mg/L at 3 months on canakinumab (threshold achievers) to complete placebo group. Does this analysis estimate a causal effect?

- YES
- NO

Assume now that we compare the threshold achievers on canakinumab to the threshold achievers on placebo. Does this analysis estimate a causal effect?

- YES
- NO

CANTOS trial

- None of these two analyses estimate a causal effect
- In both analyses treatment and population are confounded
  - Populations „threshold achievers on canakinumab “, „threshold achievers on placebo“ and „complete placebo group“ are all different
    - „threshold achievers on treatment“ likely to have lower baseline hsCRP than „complete placebo“ → also likely to have better outcomes (hsCRP is prognostic)
    - „threshold achievers on placebo“ likely to have lower baseline hsCRP than „threshold achievers on treatment“ → even more likely to have better outcomes

CANTOS trial

- Let T(1) and T(0) denote time to event and S(1), S(0) hsCRP threshold achievement under treatment and placebo
- Estimand of interest: P(T(1) > t | S(1) = 1) – P(T(0) > t | S(1) = 1)
  - Survival probability at time t in the subgroup of patients that would hsCRP threshold achievers at 3 months if on canakinumab
  - Principal stratum estimand
- One possible analysis assumption: Identifying confounders X on hsCRP response S(1) and outcome T(0) allows to re-introduce balance (e.g. using methods introduced in the next session)
  - + careful handling of competing risk situation (intercurrent event vs MACE event)

Controversies in causal inference

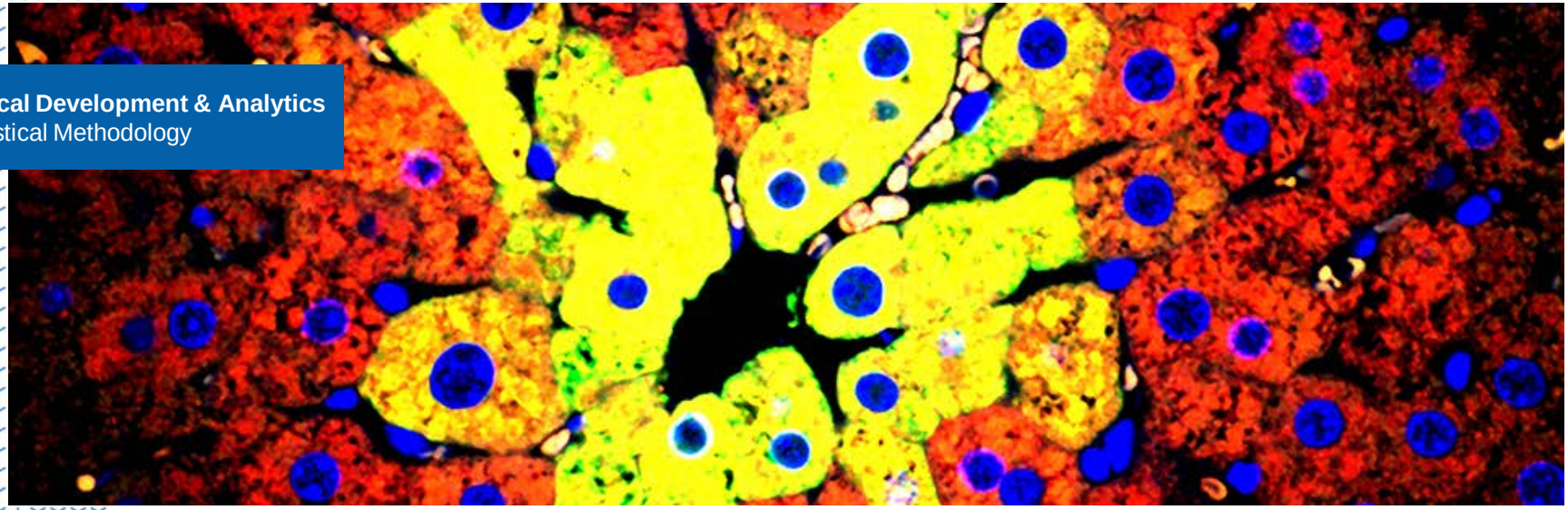
- Causal inference split into different „schools“
  - Three main figures: Donald Rubin, Jamie Robins & Judea Pearl
  - All came to causal inference from slightly different angles (with own notation etc)
- View on usefulness of DAGs
- No causation without manipulation
- Across-world assumptions
- Bayesian versus Frequentist

Why do we need causal inference?

- Provides a language to discuss causal effects (potential outcomes & DAGs)
  - applies in observational *and* randomized data situations (estimands underlying most „standard analyses“ can be described in potential outcome language)
  - See also Lipkovich et al (2020)
- Sheds new light on the understanding of some standard statistical practices
  - LS means, interpretability of treatment effect parameters (odds ratio and hazard ratio), see next session
- Will help implementing the ICH E9 addendum
  - Adopts counterfactual viewpoint to define treatment effects
  - Estimand strategies can be clearly described using potential outcome language
  - Not all estimand strategies require specialized causal inference analysis techniques

Q & A

Clinical Development & Analytics
Statistical Methodology



A Gentle Introduction to Causal Inference in View of the ICH E9 Addendum on Estimands

Björn Bornkamp, Heinz Schmidli, Dong Xi

ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop
September 22, 2020

NOVARTIS | Reimagining Medicine

Agenda

10:00 – 11:40 AM

- Introduction to causal effects & potential outcomes (Heinz Schmidli)
- Relation to questions and concepts encountered in randomized clinical trials (Björn Bornkamp)

12:00 – 1:30 PM

- Standardization & inverse probability weighting (Dong Xi)

Part 2: Standardization & inverse probability weighting

Outline

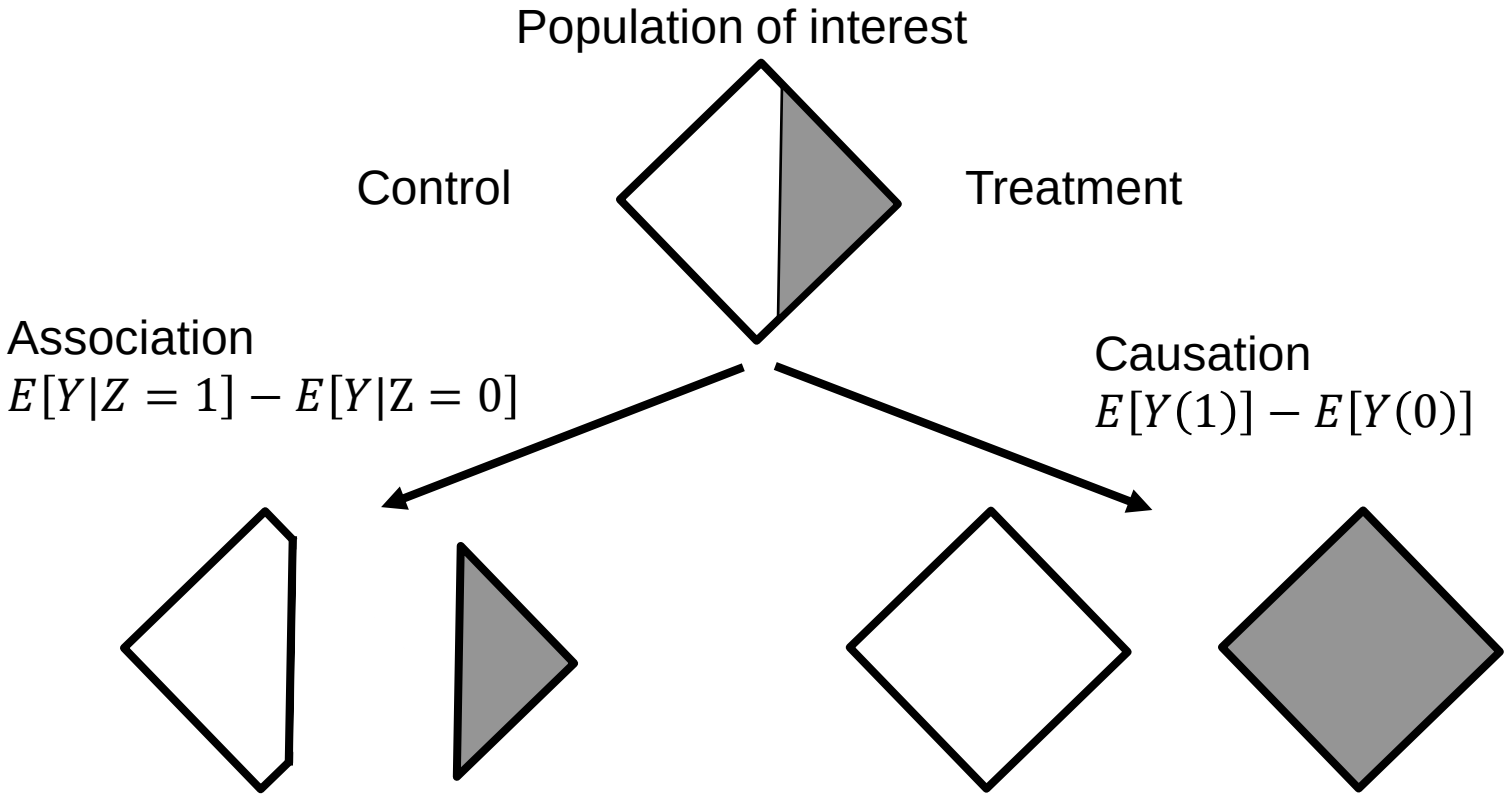
- Causal effect under (stratified) randomization
  - Standardization
  - Inverse probability weighting
- Extension to non-randomized data
- Conclusions

Fundamental problem of causal inference

- For every patient, there are two potential outcomes
  - $Y(1)$ under treatment $(Z = 1)$
  - $Y(0)$ under control $(Z = 0)$
- Only one observed outcome
  - $Y = Y(1)$ if $Z = 1$
  - $Y = Y(0)$ if $Z = 0$
  - Fundamental problem of causal inference (Holland, 1986)
- Causal inference aims to use observed outcomes (Y) to drawing conclusions about potential outcomes ($Y(1)$ and $Y(0)$)

$$E[Y(1)] - E[Y(0)] \overset{?}{=} E[Y|Z = 1] - E[Y|\mathrm{Z} = 0]$$

Association vs. causation



Magic of randomization

- Because of randomization, the treated patients are “similar” to the control patients
  - Similarity with respect to measured covariates (e.g., age, gender, weight...)
  - More importantly, with respect to unmeasured covariates and potential outcomes
- Randomization implies exchangeability: $Y(0), Y(1) \perp Z$
  - Potential outcomes are independent of (or balanced with respect to) treatment assignment
- Often use mean exchangeability (implied by exchangeability)

$$E[Y(z)|Z = 0] = E[Y(z)|Z = 1]$$

  - Mean of potential outcomes that would be observed with $Z = z$ is the same among those who actually got $Z = 0$ and those who got $Z = 1$

Causal effect under exchangeability

- Under exchangeability or mean exchangeability

$$E[Y(z)|Z=0] = E[Y(z)|Z=1] = E[Y(z)]$$

- Identify the causal effect

$$E[Y(1)] - E[Y(0)] = E[Y(1)|Z=1] - E[Y(0)|Z=0] \quad \text{(mean) exchangeability}$$

$$= E[Y|Z=1] - E[Y|Z=0] \qquad Y = \begin{cases} Y(1) \text{ if } Z=1 \\ Y(0) \text{ if } Z=0 \end{cases}$$

- Under randomization, the observed mean difference is the causal mean difference

Measurement of causal effect

- Individual causal effect $Y(1) - Y(0)$
- Average causal effect in a population is often of interest
  - $E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$
  - Population average (or marginal) effect: averaging (or marginalizing) over all individual-level effects in the population
- Other causal effect measures for binary and count data
  - Rate ratio: $E[Y(1)]/E[Y(0)]$
  - Odds ratio for binary data: $\frac{E[Y(1)]}{1-E[Y(1)]} \Big/ \frac{E[Y(0)]}{1-E[Y(0)]}$

Causal effect under conditional exchangeability

- Stratified randomization to increase homogeneity within a stratum (e.g., stratification by smokers vs. non-smokers)
  - Ensure the treated patients are “similar” to the control patients within a stratum
- Stratified randomization implies conditional exchangeability: $Y(0), Y(1) \perp Z|X$
- Conditional (mean) exchangeability within a stratum of $X = x$

$$E[Y(z)|Z = 0, X = x] = E[Y(z)|Z = 1, X = x]$$

  - Mean of potential outcomes that would be observed with $Z = z$ is the same among those who actually got $Z = 0$ and those who got $Z = 1$, within the stratum of $X = x$

Poll question 4

- Which of the following methods are you familiar with?
  - Generalized Linear Models
  - LS Mean
  - Standardization
  - Inverse Probability Weighting

Generalized Linear Models

- For a given model (e.g., GLM), there is a natural parameter of interest
  - E.g., regression parameter β of the treatment assignment
- We know there are statistical estimation techniques (e.g., maximum likelihood) to derive an unbiased estimator of the parameter in the model
- But we do not know if the parameter in the model is the parameter of interest (or the summary measure) in the estimand

> *An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question.*
>
> *John Tukey*

Poll question 5

- Given the model fit: $\hat{E}[Y|Z,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 Z + \hat{\beta}_2 X\}$, it is natural to use the regression coefficient $\hat{\beta}_1$ as an estimator of the estimand
  - $\hat{\beta}_1$ is usually called the conditional effect, conditioning on the covariate(s)
  - Steingrimsson et al. (2017)

- Which of the following models provide(s) a conditional effect ($\hat{\beta}_1$) that coincides with the population average effect below (or target the estimand of interest)?
  - Linear regression for $E[Y(1)] - E[Y(0)]$
  - Logistic regression for $\frac{E[Y(1)]}{1-E[Y(1)]} / \frac{E[Y(0)]}{1-E[Y(0)]}$
  - Poisson / Negative binomial regression for $E[Y(1)]/E[Y(0)]$

Poll question 5

- Given the model fit: $\hat{E}[Y|Z,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 Z + \hat{\beta}_2 X\}$, it is natural to use the regression coefficient $\hat{\beta}_1$ as an estimator of the estimand
  - $\hat{\beta}_1$ is usually called the conditional effect, conditioning on the covariate(s)
  - Steingrimsson et al., 2017
- Which of the following models provide(s) a conditional effect ($\hat{\beta}_1$) that coincides with the population average effect below (or target the estimand of interest)?
  - Linear regression for $E[Y(1)] - E[Y(0)]$: Yes
  - Logistic regression for $\frac{E[Y(1)]}{1-E[Y(1)]} / \frac{E[Y(0)]}{1-E[Y(0)]}$: No
  - Poisson / Negative binomial regression for $E[Y(1)]/E[Y(0)]$: Yes
- Different answers depend on the functional form of the link function
  - See the following three slides for detailed explanation

Estimand of linear regression coefficient

- Linear regression: $g^{-1}\{\cdot\} = \cdot$

$$\hat{E}[Y(1)] - \hat{E}[Y(0)] = \frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(1)|X = x_i] - \frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(0)|X = x_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_2 x_i\right)$$

$$= \hat{\beta}_1 \text{ (Yes)}$$

Estimand of logistic regression coefficient

- Logistic regression: $g^{-1}\{\cdot\} = \frac{\exp(\cdot)}{\exp(\cdot)+1} \equiv \text{expit}(\cdot)$

$$\frac{\hat{E}[Y(1)]}{1-\hat{E}[Y(1)]} \Big/ \frac{\hat{E}[Y(0)]}{1-\hat{E}[Y(0)]} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(1)|X=x_i]}{1-\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(1)|X=x_i]} \Big/ \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(0)|X=x_i]}{1-\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(0)|X=x_i]}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2 x_i)}{1-\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2 x_i)} \Big/ \frac{\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_2 x_i)}{1-\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_2 x_i)}$$

$$\neq \exp\hat{\beta}_1 \text{ (No)}$$

- Note that if we plug in the mean of covariate X

$$\frac{\hat{E}[Y(1)|X=\bar{x}]}{1-\hat{E}[Y(1)|X=\bar{x}]} \Big/ \frac{\hat{E}[Y(0)|X=\bar{x}]}{1-\hat{E}[Y(0)|X=\bar{x}]} = \frac{\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2\bar{x})}{1-\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2\bar{x})} \Big/ \frac{\text{expit}(\hat{\beta}_0+\hat{\beta}_2\bar{x})}{1-\text{expit}(\hat{\beta}_0+\hat{\beta}_2\bar{x})}$$

$$= \frac{\exp(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2\bar{x})}{\exp(\hat{\beta}_0+\hat{\beta}_2\bar{x})}$$

$$= \exp\hat{\beta}_1 \text{ (i.e., conditional effect on the mean of covariate)}$$

Estimands of Poisson/negative binomial coefficient

- Poisson / Negative binomial regression: $g^{-1}\{\cdot\} = \exp(\cdot)$

$$\frac{\hat{E}[Y(1)]}{\hat{E}[Y(0)]} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(1)|X=x_i]}{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(0)|X=x_i]}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\exp(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2 x_i)}{\frac{1}{n}\sum_{i=1}^{n}\exp(\hat{\beta}_0+\hat{\beta}_2 x_i)}$$

$$= \exp\hat{\beta}_1 \text{ (Yes, under the following assumptions)}$$

- Note that $\frac{\hat{E}[Y(1)]}{\hat{E}[Y(0)]}$ is the ratio of rates assuming every patient would have the same exposure (or offset)
- Also assume no Z by X interactions

Regulatory feedback on estimate of causal effect

- A PhIII clinical trial comparing treatment against control
- Primary estimand uses the marginal (population average) odds ratio $\frac{p_1}{1-p_1}/\frac{p_0}{1-p_0}$
  - p_1 and p_0 are response rates in treatment and control arms, respectively
- Primary analysis uses the logistic regression with covariates
  - Regression coefficient as the estimate of the primary estimand

FDA: Estimand uses the marginal odds ratio but the logistic regression uses the conditional odds ratio, which does not align with the estimand

Poll question 6

- Given the model fit: $\hat{E}[Y|Z,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 Z + \hat{\beta}_2 X\}$, it is natural to use the least squares mean as an estimate of the marginal mean $E[Y(z)]$
  - LS mean plugs in the average of covariates $\hat{E}[Y|z,X=\bar{x}] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 z + \hat{\beta}_2 \bar{x}\}$
  - LS mean estimates the effect of a “special” patient with average values of covariates

- Which of the following models provide(s) an LS mean $\hat{E}[Y|z,X=\bar{x}]$ that coincides with the marginal mean $\hat{E}[Y(z)]$?
  - Linear regression
  - Logistic regression
  - Poisson / Negative binomial regression

Poll question 6

- Given the model fit: $\hat{E}[Y|Z,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 Z + \hat{\beta}_2 X\}$, it is natural to use the least squares mean as an estimate of the marginal mean $E[Y(z)]$
  - LS mean plugs in the average of covariates $\hat{E}[Y|z, X = \bar{x}] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 z + \hat{\beta}_2 \bar{x}\}$
  - LS mean estimates the effect of a "special" patient with average values of covariates
- Which of the following models provide(s) an LS mean $\hat{E}[Y|z, X = \bar{x}]$ that coincides with the marginal mean $\hat{E}[Y(z)]$?
  - Linear regression: Yes
  - Logistic regression: No
  - Poisson / Negative binomial regression: No
- Again, different answers depend on the functional form of the link function
  - See the following three slides for detailed explanation

Estimand of linear regression LS mean

- Linear regression: $g^{-1}\{\cdot\} = \cdot$

$$\hat{E}[Y(z)] = \frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(z)|X = x_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_i\right)$$

$$= \hat{\beta}_0 + \hat{\beta}_1 z + \hat{\beta}_2 \bar{x}$$

$$= \hat{E}[Y|z, X = \bar{x}] \text{ (Yes)}$$

Estimand of logistic regression LS mean

- Logistic regression: $g^{-1}\{\cdot\} = \frac{\exp(\cdot)}{\exp(\cdot)+1} \equiv \text{expit}(\cdot)$

$$\hat{E}[Y(z)] = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(z)|X=x_i]}{1-\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(z)|X=x_i]}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1 z+\hat{\beta}_2 x_i)}{1-\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1 z+\hat{\beta}_2 x_i)}$$

$$\neq \text{expit}(\hat{\beta}_0+\hat{\beta}_1 z+\hat{\beta}_2 \bar{x})$$

$$= \hat{E}[Y|z, X=\bar{x}] \text{ (No)}$$

- Interpretation: LS mean estimates the effect of a “special” patient with average values of covariates

Estimands of Poisson/negative binomial regression LS mean

- Poisson / Negative binomial regression: $g^{-1}\{\cdot\} = \exp(\cdot)$

$$\hat{E}[Y(z)] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y(1)|X = x_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \exp\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_i\right)$$

$$\neq \exp\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 \bar{x}\right)$$

$$= \hat{E}[Y|z, X = \bar{x}] \text{ (No)}$$

- Interpretation: LS mean estimates the effect of a “special” patient with average values of covariates

Regulatory feedback on marginal mean

- SIROCCO, a PhIII trial to compare benralizumab to placebo for severe asthma
- Primary analysis uses standardization from negative binomial regression of the number of exacerbations with covariates

> *FDA: The study SAP proposed the marginal standardization method in calculating mean annual exacerbation rates.*

> *FDA: We agree with the applicant's proposal in that, in the negative binomial regression setting, the marginal method more closely aligns with the crude annual exacerbation rate, and as such, provides a more appropriate covariate-adjusted summary within treatment groups.*

https://www.fda.gov/media/110333/download

Standardization

- Regression analysis is a general approach to analyze randomized/non-randomized data by adjusting for
  - Stratification variables under stratified randomization
  - Other categorical covariates to address chance imbalance
  - Other covariates for efficiency of estimation
- How to find a valid estimate of the average causal effect given these covariate adjustments?
- Standardization (standardized estimator) is a popular approach
  1. Model fitting
  2. Predicting
  3. Averaging

Step 1 in standardization: Fit a regression model (e.g., GLM)

Treatment (Z)	Covariates (X)	Response (Y)
1	x_1	y_1
0	x_2	y_2
⋮	⋮	⋮

Regress Y over Z and X

Model fit: $\hat{E}[Y|Z,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 Z + \hat{\beta}_2 X\}$

- $g\{\cdot\}$: the link function
- Linear regression: identity link $g\{\cdot\} = \cdot$
- Poisson or negative binomial regression: log link $g\{\cdot\} = \log(\cdot)$
- Logistic regression: logit link $g\{\cdot\} = \log(\frac{\cdot}{1-\cdot})$

Step 2 in standardization: Predict potential outcomes

All patients under $z = 0$

Treatment $z = 0$	Covariate (X)
0	x_1
0	x_2
⋮	⋮

Model fit: $\hat{E}[Y|Z,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 Z + \hat{\beta}_2 X\}$

Predict

All patient under $z = 1$

Treatment $z = 1$	Covariate (X)
1	x_1
1	x_2
⋮	⋮

Potential response under $z = 0$
$\hat{E}[Y(0)\|X = x_1] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_2 x_1\}$
$\hat{E}[Y(0)\|X = x_2]$
⋮

Potential response under $z = 1$
$\hat{E}[Y(1)\|X = x_1] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_1\}$
$\hat{E}[Y(1)\|X = x_2]$
⋮

Step 3 in standardization: Average over individual predictions

Potential response under $z = 0$
$\hat{E}[Y(0)\|X = x_1] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_2 x_1\}$
$\hat{E}[Y(0)\|X = x_2]$
$\vdots$

Potential response under $z = 1$
$\hat{E}[Y(1)\|X = x_1] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_1\}$
$\hat{E}[Y(1)\|X = x_2]$
$\vdots$

Averaging (marginalizing over covariates)

$$\hat{E}[Y(0)] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y(0)|X = x_i]$$

Averaging (marginalizing over covariates)

$$\hat{E}[Y(1)] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y(1)|X = x_i]$$

Estimated population average causal treatment effect

$$\hat{E}[Y(1)] - \hat{E}[Y(0)]$$

Implementation of standardization

- SAS macro “Margins” fits the GLM or GEE model and estimates marginal mean and population average treatment effects (i.e., difference in means)
  - Compatible with GENMOD
  - Use the delta method for confidence intervals, p-values
  - https://support.sas.com/kb/63/038.html
- A general approach using bootstrap
  - Create bootstrap datasets using SURVEYSELECT in SAS or boot in R
  - Within each dataset, complete steps 1 (model fitting), 2 (predicting), 3 (averaging)
  - Summarize over the bootstrap datasets for confidence intervals, p-values

Application

- A PhIII trial comparing treatment against control
- Primary endpoint
  - Clinical responder ($Y = 1$) or non-responder ($Y = 0$)
  - Logistic regression of Y on treatment and covariates
- Compare standardization vs conditional results from logistic regression
  - Odds ratio
  - Difference in probabilities

Marginal mean

Marginal mean	Method based on logistic regression	Mean (95% CI)
Control	LS mean	0.19 (0.16, 0.22)
	Standardization via Margins macro	0.23 (0.20, 0.26)
	Standardization via bootstrap	0.23 (0.20, 0.26)
Treatment	LS mean	0.51 (0.46, 0.56)
	Standardization via Margins macro	0.56 (0.52, 0.61)
	Standardization via bootstrap	0.56 (0.52, 0.61)

- Very close results between Margins macro and bootstrap
- Different results from LS mean estimates

Treatment effect

Treatment effect	Method based on logistic regression	Mean (95% CI)	P-value
Difference	Model estimate	NA	
	Standardization via Margins macro	0.33 (0.28, 0.39)	<0.001
	Standardization via bootstrap	0.33 (0.28, 0.39)	<0.001
Odds ratio	Model estimate	4.59 (3.52 5.99)	<0.001
	Standardization via Margins macro	NA	
	Standardization via bootstrap	4.40 (3.39, 5.63)	<0.001

- Very close results for difference between Margins macro and bootstrap
- Different results for odds ratio between model estimate and standardization

Properties of standardization

- Standardization derives population averages on the outcome scale
  - Coincides with the linear model estimator
  - More interpretable for discrete outcomes
  - Incorporates covariates for efficiency
- Standardization is more robust (than a regression model) to model misspecification under randomization
  - Consistent estimator even when the GLM is misspecified (e.g., wrong choice of covariates)
- Standardization provides flexible estimators for different effect measures (difference, ratio, odds ratio etc.)
- More awareness on what estimands are targeted by common estimators
  - Rosenblum and van der Laan (Int J Biostat, 2010) for GLM
  - Hernán (Epidemiology, 2010) for hazard ratio

Q & A

When data or comparisons are not randomized

- Real-world data/evidence are often observational in nature
- Even in randomized trials, intercurrent events may lead to comparisons of two post-randomized groups
  - E.g., protocol adherers may be different under treatment and control
  - E.g., treatment switching is often based on patients' post-randomized condition
- Without randomization, (conditional) exchangeability may not hold

$$?\, Y(0), Y(1) \perp Z|X \,?$$

- To address this, we need to make additional assumptions
  - No unmeasured confounding
  - Positivity

No unmeasured confounding

- Because of randomization, treated patients are “similar” to the control patients
  - Similarity with respect to measured & unmeasured covariates
  - And therefore also potential outcomes
- Without randomization, we need to assume that we have measured all possible covariates that affect both treatment assignment and outcome
  - i.e. no unmeasured confounders
- Given this assumption, (conditional) exchangeability holds with “appropriate” $\boldsymbol{X}$

$$Y(0), Y(1) \perp Z|\boldsymbol{X}$$

- Given $\boldsymbol{X}$ (within strata of $\boldsymbol{X}$), we believe the treatment assignment is “random”

Positivity

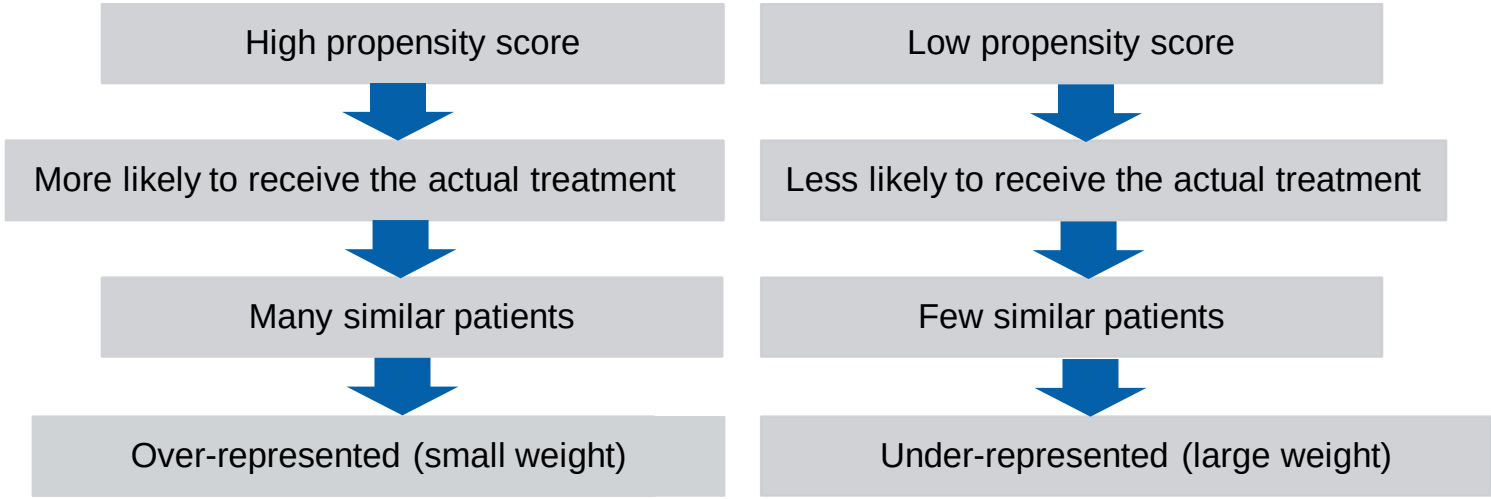
- Positivity: $P(Z = z|\boldsymbol{X} = \boldsymbol{x}) > 0$ for everyone
  - Patients always had the possibility to receive (or not) any treatment option
- Why needed?
  - A patient with $P(Z = 1) = 1$ implies there is no comparable patient who did not receive treatment
  - i.e. we know nothing about their potential outcome under the treatment they did not receive
- Randomization implies positivity

Standardization in non-randomized studies

- Standardization can still be used as described in the previous section even in non-randomized studies
  - As long as, there is no unmeasured confounding and positivity holds,
  - And the regression model includes all confounders,
  - And the model is correctly specified
- In complex studies, correct model specification ($E[Y|Z,X]$) can be difficult
- Inverse probability weighting is an alternative method that does not require the correct model specification ($E[Y|Z,X]$)
  - But requires the correct model for weights

Inverse probability weighting (IPW)

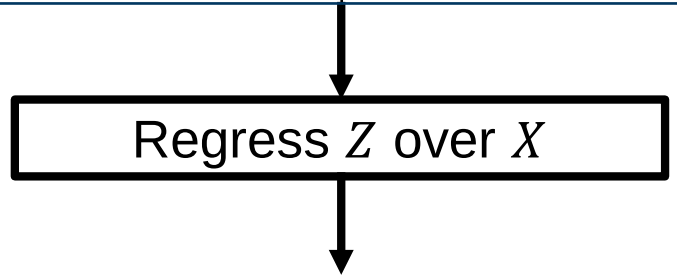
- Model the treatment assignment $P(Z = z|\boldsymbol{X} = \boldsymbol{x})$, i.e., propensity score



- Weight inversely proportional to the propensity score $\propto 1/P(Z = z|\boldsymbol{X} = \boldsymbol{x})$

Step 1 in IPW:
Fit a propensity score model (e.g., logistic)

Treatment (Z)	Covariates (X)
1	x_1
0	x_2
⋮	⋮



Model fit: $\hat{P}(Z = 1|X) = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_2 X\}$

Step 2 in IPW: Predict propensity score

Treatment (Z)	Covariates (X)
1	x_1
0	x_2
⋮	⋮

Model fit: $\hat{P}(Z = 1|\boldsymbol{X}) = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_2 X\}$

Predict

Treatment (Z)	Covariates (X)	$\widehat{P}(Z = 1|X)$	$\widehat{P}(Z = z|X)$
1	x_1	p_1	p_1
0	x_2	p_2	$1 - p_2$
⋮	⋮	⋮	⋮

Step 3 in IPW: Weighted regression of outcome Y

Treatment (Z)	Covariates (X)	$\hat{P}(Z = z\|X)$	Weight
1	x_1	p_1	$\frac{\hat{P}(Z=1)}{p_1}$
0	x_2	$1 - p_2$	$\frac{\hat{P}(Z=0)}{1-p_2}$
⋮	⋮	⋮	

$\hat{P}(Z = z)$: proportion of patients who received z

Regress Y over Z, using $\hat{P}(Z = z)/\hat{P}(Z = z|X)$ as weights

See slide 94 for other weights

Application

- A PhIII trial comparing treatment against control
- Primary endpoint
  - Clinical responder ($Y = 1$) or non-responder ($Y = 0$)
  - Logistic regression of Y on treatment and covariates
- Propensity score model
  - Logistic regression of Z on covariates
- Analysis model for IPW
  - Weighted logistic regression of Y on treatment only
  - Close to taking weighted means in each treatment group

Marginal mean

Marginal mean	Method based on logistic regression	Mean (95% CI)
Control	LS mean	0.19 (0.16, 0.22)
	IPW	0.23 (0.20, 0.26)
	Standardization	0.23 (0.20, 0.26)
Treatment	LS mean	0.51 (0.46, 0.56)
	IPW	0.58 (0.54, 0.63)
	Standardization	0.56 (0.52, 0.61)

- Similar results between IPW and standardization
- Different results from LS mean estimates

Treatment effect

Treatment effect	Method based on logistic regression	Mean (95% CI)	P-value
Difference	Model estimate	NA	
	IPW	0.35 (0.30, 0.41)	<0.001
	Standardization	0.33 (0.28, 0.39)	<0.001
Odds ratio	Model estimate	4.59 (3.52, 5.99)	<0.001
	IPW	4.68 (3.60, 6.12)	<0.001
	Standardization	4.40 (3.39, 5.63)	<0.001

- Similar results for difference between IPW and standardization
- Different results for odds ratio
  - Odds ratio is very sensitive to a small change in the marginal mean

Properties of IPW

- Propensity score is a balancing score
  - Independence between treatment and covariates given propensity score
  - If conditional exchangeability holds for covariates X, it holds for propensity score $P(Z = z|X)$
- Propensity score model should include all possible (baseline) variables that could affect treatment and outcome
  - More important to derive weights that improve covariate balance than to predict treatment
- IPW targets the population average parameters, under the correct model
- Bootstrap is generally valid for inference of IPW
  - Usually, **stabilized weights are preferred**: $\Pr(Z = z) / \Pr(Z = z|X)$
  - Or truncated weights
- Propensity score has a broad use in weighting (for treatment, censoring etc.), matching and stratification
  - Need to check for positivity (overlap in propensity score)

Conclusion

- Randomization allows identification of causal effects from observed data
- Without randomization or in the presence of post-randomized comparisons, assumptions are needed to mimic a randomized setting
- Standardization is a robust and efficient approach under randomization
- With and without randomization
  - Both standardization and IPW target the population average causal effect (or estimand)
  - Their estimates could be different due to the used of different statistical models
  - Standardization relies on an outcome model with covariates
  - IPW models treatment assignment via a propensity score model with covariates
  - Doubly-robust methods combine standardization and IPW and thus are more robust to model misspecification
- Important to understand properties of estimators for a better alignment with estimands

Q & A

Why do we need causal inference?

- Provides a language to discuss causal effects (potential outcomes & DAGs)
  - applies in observational and randomized data situations
- Understand properties of estimators for a better alignment with estimands
  - What estimand is the chosen estimator targeting?
  - What are the assumptions underlying the estimator and how plausible are they?
- Sheds new light on the understanding of some standard statistical practices
  - LS means, interpretability of treatment effect parameters
  - Some current standard practices might change
- Will help implementing the ICH E9 addendum
  - Adopts counterfactual viewpoint to define treatment effects
  - Causal thinking & techniques apply to all intercurrent event strategies

References

- Bornkamp, B, Bermann, G (2020). Estimating the treatment effect in a subgroup defined by an early post-baseline biomarker measurement in randomized clinical trials with time-to-event endpoint. *Stat Biopharm Res.*
- Ge M et al. (2011). Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. *Drug Inf J.*
- Hernán MA (2010). The hazards of hazard ratios. *Epidemiology*.
- Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC, forthcoming www.hsph.harvard.edu/miguel-hernan/causal-inference-book
- Imbens GW, Rubin DB (2015) Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.
- Lipkovich I, Ratitch B, Mallinckrodt CH (2020). Causal inference and estimands in clinical trials. *Stat Biopharm Res*.
- Lou Y, Jones MP, Sun W (2019). Estimation of causal effects in clinical endpoint bioequivalence studies in the presence of intercurrent events: noncompliance and missing data. *J Biopharm Stat*.
- Pearl J, Mackenzie D (2018) The Book of Why: The New Science of Cause and Effect. Hachette.
- Ridker P. et al (2017) Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease, *New England Journal of Medicine*; 377:1119-1131
- Rosenblum M, Van Der Laan MJ (2010). Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *Int J Biostat*.
- Steingrimsson JA, Hanley DF, Rosenblum M (2017). Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemp Clin Trials*.

Thank you

NOVARTIS | Reimagining Medicine