Covariate Adjustment

SISCER, July 2024

Kelly Van Lancker

UNIVERSITEIT GENT

FACULTY OF SCIENCES

Big thank you to

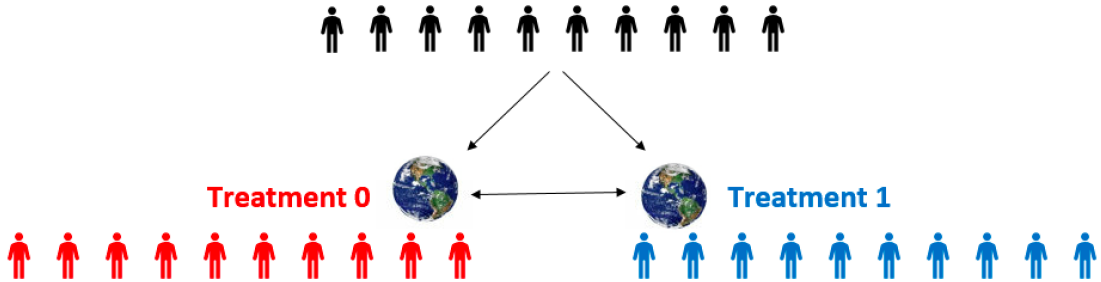
- Michael Rosenblum and Josh Betz (Johns Hopkins University)
- Frank Bretz (Novartis)
- Stijn Vansteelandt and Oliver Dukes (Ghent University)

Outline

Potential Outcomes

- Consider an eligible patient population.

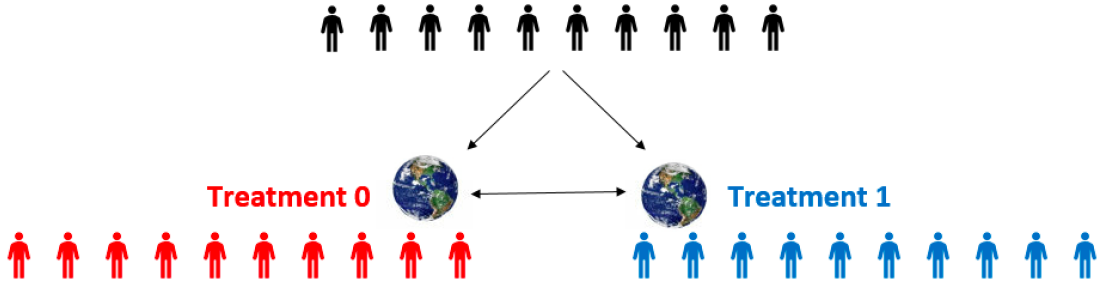
Potential Outcomes



- Consider an eligible patient population.
- Imagine two parallel worlds: one where everyone is assigned Treatment 0 and one where everyone is assigned Treatment 1.
  - Y^0 and Y^1: **potential/hypothetical outcomes** in the two parallel worlds.
  - Superscript (0 or 1): **allocation to treatment**.

Outline

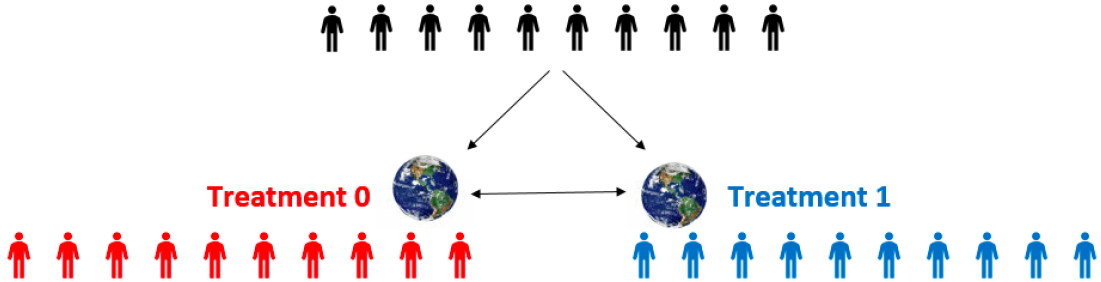
Hypothetical World: Causal Estimand



Causal Estimand

Average of the outcomes when everyone is assigned to Treatment 1
minus
average of the outcomes when everyone is assigned to Treatment 0

Hypothetical World: Causal Estimand



Causal Estimand

Average of the outcomes when everyone is assigned to Treatment 1
minus
average of the outcomes when everyone is assigned to Treatment 0

Mean difference: $E\left(Y^1\right) - E\left(Y^0\right)$

Marginal Causal Contrasts

- Causal contrasts of interest often reflect a contrast between the means of the distributions of Y^0 and Y^1: $E\left(Y^0\right)$ and $E\left(Y^1\right)$
  - Mean difference $E\left(Y^1\right) - E\left(Y^0\right)$
  - Mean ratio $E\left(Y^1\right) / E\left(Y^0\right)$
  - Odds ratio $\frac{E(Y^1)/\{1-E(Y^1)\}}{E(Y^0)/\{1-E(Y^0)\}}$
  - ...

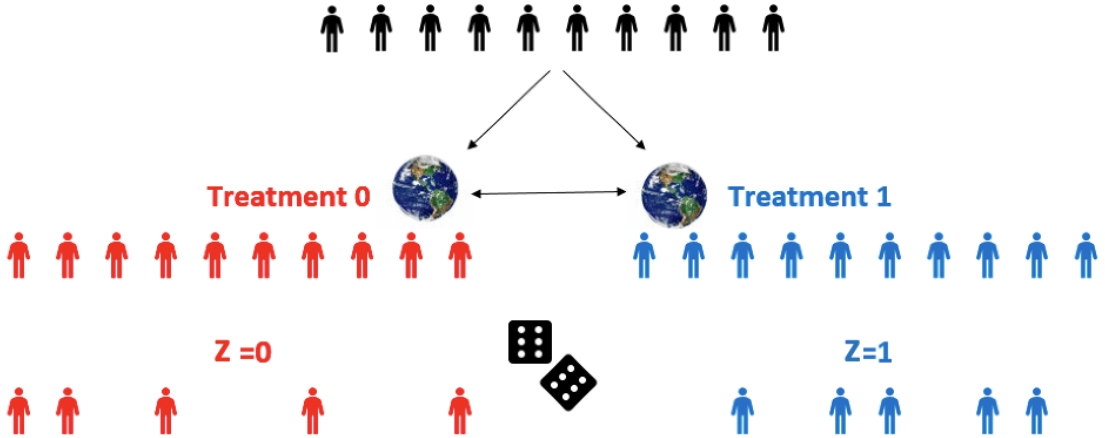
Marginal Causal Contrasts

- Causal contrasts of interest often reflect a contrast between the means of the distributions of Y^0 and Y^1: $E\left(Y^0\right)$ and $E\left(Y^1\right)$
  - Mean difference $E\left(Y^1\right) - E\left(Y^0\right)$
  - Mean ratio $E\left(Y^1\right) / E\left(Y^0\right)$
  - Odds ratio $\frac{E(Y^1)/\{1-E(Y^1)\}}{E(Y^0)/\{1-E(Y^0)\}}$
  - ...
- These are **marginal** causal contrasts.

Marginal Causal Contrasts

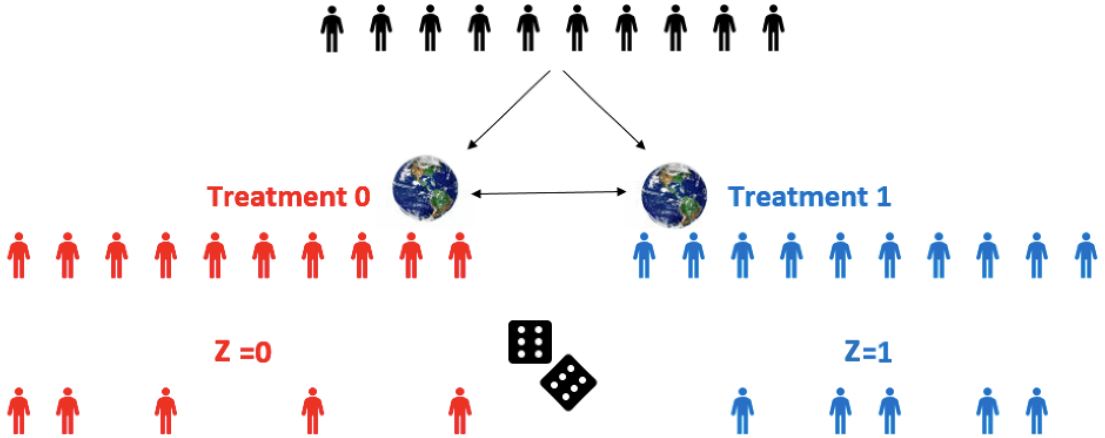
- Causal contrasts of interest often reflect a contrast between the means of the distributions of Y^0 and Y^1:
  $E\left(Y^0\right)$ and $E\left(Y^1\right)$
  - Mean difference $E\left(Y^1\right) - E\left(Y^0\right)$
  - Mean ratio $E\left(Y^1\right) / E\left(Y^0\right)$
  - Odds ratio $\frac{E(Y^1)/\{1-E(Y^1)\}}{E(Y^0)/\{1-E(Y^0)\}}$
  - ...
- These are **marginal** causal contrasts.
- The (marginal) causal contrast can also be a contrast of other summaries of the distributions of Y^0 and Y^1;
  e.g., for time-to-event outcomes.

Real world: Randomization



- In real life, patients are randomized to only one group.

Real world: Randomization

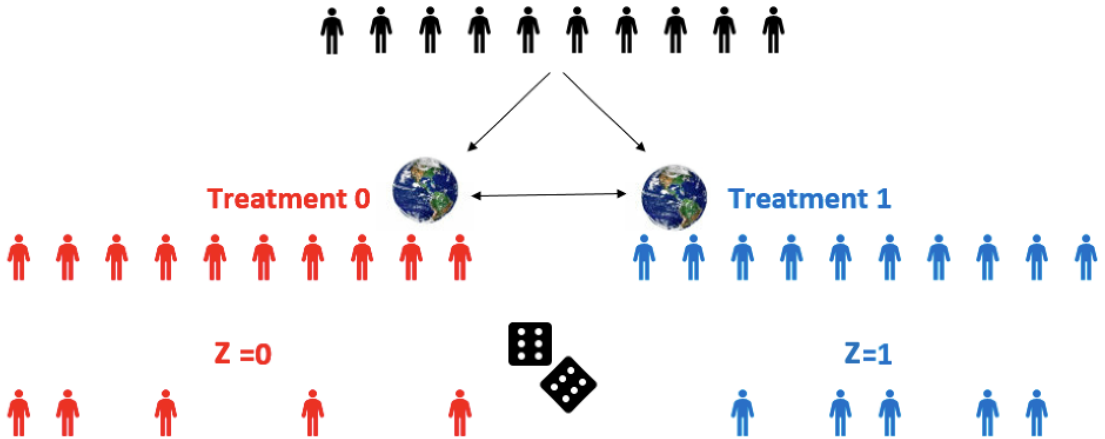


- In real life, patients are randomized to only one group.

Causal Treatment Effect Estimate

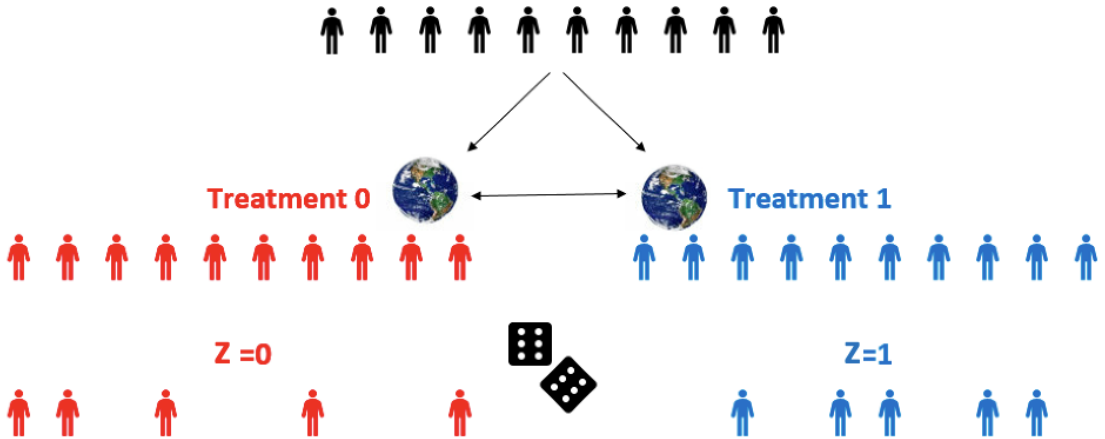
Average of **observed** *outcomes of patients assigned to Treat. 1*
minus
average of **observed** *outcomes of patients assigned to Treat. 0*

Randomization



- In real life, patients are randomized to only one group.
  - The randomized group is denoted Z and the factual/observed outcome Y.

Randomization



- In real life, patients are randomized to only one group.
  - The randomized group is denoted Z and the factual/observed outcome Y.
- Randomization ensures that causal contrasts correspond to statistical contrasts:
  - $E\left(Y^1\right) - E\left(Y^0\right) = E\left(Y|Z=1\right) - E\left(Y|Z=0\right)$.

Outline

Conditional Causal Contrasts

- So far, we have been focusing on **marginal estimands**.
  - A (causal) treatment effect for the **whole eligible patient population**.

Conditional Causal Contrasts

- So far, we have been focusing on **marginal estimands**.
  - A (causal) treatment effect for the **whole eligible patient population**.

- We can also focus on a certain subgroup

Conditional Causal Contrasts

- We can also focus on a certain subgroup
  - For example, there may be interest in the treatment effect (on a certain scale) in the male or female participants separately.

Conditional Causal Contrasts

- We can also focus on a certain subgroup
  - For example, there may be interest in the treatment effect (on a certain scale) in the male or female participants separately.
  - These are **conditional** (i.e., within stratum of baseline variable(s)) treatment effects.

Conditional Causal Contrasts

- We can also focus on a certain subgroup
  - For example, there may be interest in the treatment effect (on a certain scale) in the male or female participants separately.
  - These are **conditional** (i.e., within stratum of baseline variable(s)) treatment effects.
  - We can just take the difference in means between the outcomes of female/male participants under Treatment 1 and Treatment 0.

Conditional Causal Contrasts

- Thus, causal contrasts of interest can also reflect a contrast between the means of the distributions of Y^0 and Y^1 **in a subset of patients (e.g., females)**:
  - e.g., mean difference $E\left(Y^1|sex = f\right) - E\left(Y^0|sex = f\right)$

Conditional Causal Contrasts

- Thus, causal contrasts of interest can also reflect a contrast between the means of the distributions of Y^0 and Y^1 **in a subset of patients (e.g., females)**:
  - e.g., mean difference $E\left(Y^1|sex = f\right) - E\left(Y^0|sex = f\right)$

- Randomization ensures that

$$E\left(Y^1|sex = f\right) - E\left(Y^0|sex = f\right) = E\left(Y|Z = 1, sex = f\right) - E\left(Y|Z = 0, sex = f\right).$$

Conditional Causal Contrasts

- Thus, causal contrasts of interest can also reflect a contrast between the means of the distributions of Y^0 and Y^1 **in a subset of patients (e.g., females)**:
  - e.g., mean difference $E\left(Y^1|sex = f\right) - E\left(Y^0|sex = f\right)$
- Randomization ensures that

  $E\left(Y^1|sex = f\right) - E\left(Y^0|sex = f\right) = E\left(Y|Z = 1, sex = f\right) - E\left(Y|Z = 0, sex = f\right).$
- However, estimation typically requires **model assumptions** (such as logistic regression model),
  and the estimate is often **uninterpretable** under model misspecification.

Conditional Causal Contrasts

- So far, our estimand definitions have been **model-free**.

Conditional Causal Contrasts

- So far, our estimand definitions have been **model-free**.
- ICH E9 (FDA and EMA, 1998) and EMA (2015) guidelines are written with the understanding that the **target treatment effect is a model parameter**; e.g.,

$$g\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X$$

where $g(\cdot)$ is a pre-specified link function.

Conditional Causal Contrasts

- So far, our estimand definitions have been **model-free**.
- ICH E9 (FDA and EMA, 1998) and EMA (2015) guidelines are written with the understanding that the **target treatment effect is a model parameter**; e.g.,

$$g\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X$$

  where $g(\cdot)$ is a pre-specified link function.

- This model implies the same treatment effect in the subgroups:
  it makes the **statistical modelling assumption** that there is no interaction between Z and X (on the considered scale)
  - **Not implied by randomization**.

Conditional Causal Contrasts: Continuous Outcome

- Let's consider a continuous outcome.

Conditional Causal Contrasts: Continuous Outcome

- Let's consider a continuous outcome.
- One may then choose $g(\cdot)$ to be the identity link function, and fit the model

$$E(Y|Z,X) = \beta_0 + \beta_1 Z + \beta_2 X.$$

- **Statistical modelling assumption**: no interaction between Z and X on the linear scale
  - **Not implied by randomization**.

Conditional Causal Contrasts: Continuous Outcome

- Let's consider a continuous outcome.
- One may then choose $g(\cdot)$ to be the identity link function, and fit the model

$$E(Y|Z,X) = \beta_0 + \beta_1 Z + \beta_2 X.$$

- **Statistical modelling assumption**: no interaction between Z and X on the linear scale
  - **Not implied by randomization**.
- If assumption holds:
  - β_1 carries an interpretation as *both* a conditional causal effect $E(Y^1 - Y^0|X = x)$ and a marginal causal effect $E(Y^1 - Y^0)$.

Conditional Causal Contrasts: Continuous Outcome

- One might also fit a model with an interaction

$$E(Y|Z,X) = \beta_0 + \beta_1 Z + \beta_2 X + \beta_3 ZX$$

Conditional Causal Contrasts: Continuous Outcome

- One might also fit a model with an interaction

$$E(Y|Z,X) = \beta_0 + \beta_1 Z + \beta_2 X + \beta_3 ZX$$

  - β_1: (conditional) effect in those with $X = 0$.
  - $\beta_1 + \beta_3 x$: (conditional) effect in those with $X = x$.

Conditional Causal Contrasts: Continuous Outcome

- One might also fit a model with an interaction

$$E(Y|Z,X) = \beta_0 + \beta_1 Z + \beta_2 X + \beta_3 ZX$$

  - β_1: (conditional) effect in those with $X = 0$.
  - $\beta_1 + \beta_3 x$: (conditional) effect in those with $X = x$.

- β_1 and β_3 typically lose their marginal interpretation
  - unless X is appropriately scaled (Ye et al., 2022).

Conditional Causal Contrasts: Continuous Outcome

- One might also fit a model with an interaction

$$E(Y|Z,X) = \beta_0 + \beta_1 Z + \beta_2 X + \beta_3 ZX$$

  - β_1: (conditional) effect in those with $X = 0$.
  - $\beta_1 + \beta_3 x$: (conditional) effect in those with $X = x$.

- β_1 and β_3 typically lose their marginal interpretation
  - unless X is appropriately scaled (Ye et al., 2022).

- However, we can use these models to obtain marginal treatment effect estimates by averaging across the empirical distribution of baseline covariates (see later).

Conditional Causal Contrasts: Other Outcomes

- For a binary outcome Y, it is more common to choose the logistic regression model

$$logit\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X.$$

- If the model reflects the truth, then the effect of treatment (β_1) does not differ for different values of X.

Conditional Causal Contrasts: Other Outcomes

- For a binary outcome Y, it is more common to choose the logistic regression model

$$logit\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X.$$

- If the model reflects the truth, then the effect of treatment (β_1) does not differ for different values of X.

- Unlike in the linear case, $exp(\beta_1)$ would *only* retain an interpretation as a conditional effect,

$$\frac{E(Y^1|X=x)/\{1-E(Y^1|X=x)\}}{E(Y^0|X=x)/\{1-E(Y^0|X=x)\}},$$

which may differ from the marginal causal odds ratio

$$\frac{E(Y^1)/\{1-E(Y^1)\}}{E(Y^0)/\{1-E(Y^0)\}}.$$

Non-Collapsibility

- **Standard practice based on logistic regression does not typically target a marginal effect.**

Non-Collapsibility

- **Standard practice based on logistic regression does not typically target a marginal effect.**
- This phenomenon occurs due to the **non-collapsibility** of the logistic link function; see Daniel et al. (2021).
  - Not unique to logistic regression; e.g., Cox proportional hazards models.

Non-Collapsibility

- An effect measure is **collapsible** if conditional and marginal estimands are the same: e.g., mean difference

Non-Collapsibility

- An effect measure is **collapsible** if conditional and marginal estimands are the same: e.g., mean difference
- Examples of non-collapsible effect measures:

Non-Collapsibility

- An effect measure is **collapsible** if conditional and marginal estimands are the same: e.g., mean difference
- Examples of non-collapsible effect measures:
  - the marginal **odds ratio** is not the same as the conditional odds ratio.

Non-Collapsibility

- An effect measure is **collapsible** if conditional and marginal estimands are the same: e.g., mean difference
- Examples of non-collapsible effect measures:
  - the marginal **odds ratio** is not the same as the conditional odds ratio.
  - the marginal **hazard ratio** is not the same as the conditional hazard ratio.

Illustration of non-collapsibility: odds ratio

	Males		Females		Males + Females	
	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14
Odds ratio:	= 9		= 9		= 5.4	

The effect in females is the same as the effect in males, but the effect in females and males together is different. Astonishing!

Illustration of non-collapsibility: odds ratio

	Males		Females		Males + Females	
	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14
Odds ratio:	= 9		= 9		= 5.4	

The effect in females is the same as the effect in males, but the effect in females and males together is different. Astonishing!

Non-Collapsibility

Even when all subgroup treatment effects are identical, this subgroup-specific conditional treatment effect can differ from the marginal treatment effect.

Illustration of collapsibility: risk difference

	Males		Females		Males + Females	
	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14
Odds ratio:	= 9		= 9		= 5.4	

- Conditional risk difference:
  - Males: 0.90 - 0.50 = 0.40
  - Females: 0.50 - 0.10 = 0.40
- Marginal risk difference:
  - 0.70 - 0.30 = 0.40

Collapsibility

The marginal treatment effect is a weighted average of subgroup-specific conditional treatment effects.

Conditional Causal Contrasts: Other Outcomes

- For a binary outcome Y, it is more common to choose the logistic regression model

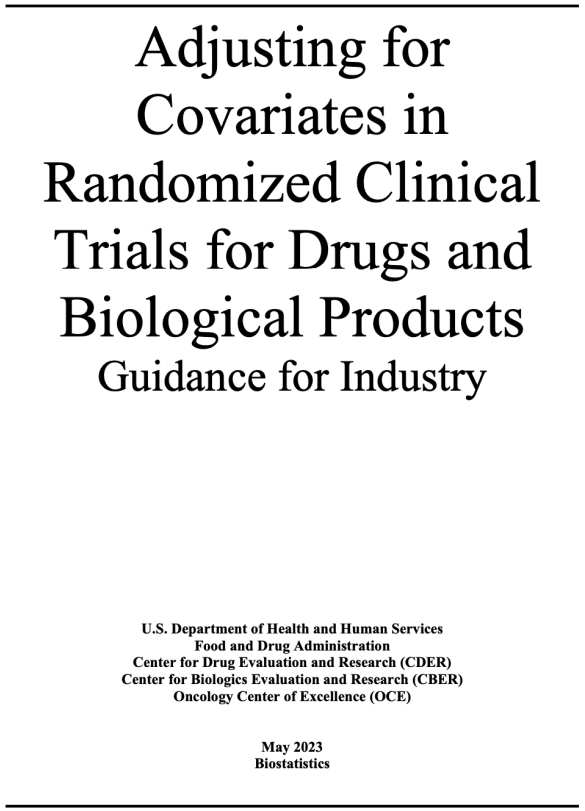
$$logit\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X.$$

- When the model is misspecified, the standard likelihood-based estimators of β_1 may not generally target either $\frac{E(Y^1|X=x)/\{1-E(Y^1|X=x)\}}{E(Y^0|X=x)/\{1-E(Y^0|X=x)\}}$ or $\frac{E(Y^1)/\{1-E(Y^1)\}}{E(Y^0)/\{1-E(Y^0)\}}$.
  - The concern for model misspecification for non-linear models is for example highlighted in the (EMA, 2015) guideline.

Outline

FDA guidance on covariate adjustment

- Choice between **marginal** and **conditional** treatment effects is an **estimand** decision.
- **Covariate adjustment** is an **analysis** decision.
  - Linear model: marginal and conditional effect estimates coincide.
  - Non-linear model: be cautious due to non-collapsibility.



Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

May 2023
Biostatistics

Conditioning versus Adjusting

- **Confusing**: Traditionally '**conditional**' and '**adjusted**' have been used interchangeably, likewise for '**marginal**' and '**unadjusted**'.

Conditioning versus Adjusting

- **Confusing**: Traditionally '**conditional**' and '**adjusted**' have been used interchangeably, likewise for '**marginal**' and '**unadjusted**'.
- Better suggestion:
  - marginal/conditional are related to the '**estimand**' of interest
  - unadjusted/adjusted are related to the '**analysis**' performed

Conditioning versus Adjusting

- **Confusing**: Traditionally '**conditional**' and '**adjusted**' have been used interchangeably, likewise for '**marginal**' and '**unadjusted**'.
- Better suggestion:
  - marginal/conditional are related to the '**estimand**' of interest
  - unadjusted/adjusted are related to the '**analysis**' performed
- Perfectly possible to obtain an adjusted estimator of a marginal estimand.
  - **Adjusted estimators of marginal estimands are almost always more precise than unadjusted estimators.**

Conditioning versus Adjusting

- **Confusing**: Traditionally '**conditional**' and '**adjusted**' have been used interchangeably, likewise for '**marginal**' and '**unadjusted**'.
- Better suggestion:
  - marginal/conditional are related to the '**estimand**' of interest
  - unadjusted/adjusted are related to the '**analysis**' performed
- Perfectly possible to obtain an adjusted estimator of a marginal estimand.
  - **Adjusted estimators of marginal estimands are almost always more precise than unadjusted estimators.**
- Recent FDA guidelines make a distinction between conditioning and adjusting (FDA, 2023).
  - Recommendations for covariate adjustment.
  - Advice on both conditional, and marginal estimands.

Covariate Adjustment for Marginal Estimands

- **Covariate adjustment** is a statistical analysis method with high potential to **improve precision** for many of these trials.
  - **Pre-planned** adjustment for baseline variables when estimating **average treatment effect**.
  - Estimand is same as when using unadjusted estimator (e.g., difference in means).
  - **Goal**: avoid making any model assumptions beyond what's assumed for unadjusted estimator (**robustness to model misspecification**).

(e.g., Koch et al., 1998; Yang and Tsiatis, 2001; Rubin and van der Laan, 2008; Tsiatis et al., 2008; Moore and van der Laan, 2009b,a; Zhang, 2015; Jiang et al., 2018; Benkeser et al., 2020)

Example

- Suppose we aim to learn the treatment effect on a binary outcome Y (e.g., 'disease').

Age	Z	Y	Y^1	Y^0
40	1	1	1	?
50	1	0	0	?
60	1	1	1	?
50	0	0	?	0
30	0	1	?	1
40	0	0	?	0

- By randomization: fine to compare outcomes of treated with outcomes of untreated

Example

- Suppose we aim to learn the treatment effect on a binary outcome Y (e.g., 'disease').

Age	Z	Y	Y^1	Y^0
40	1	1	1	?
50	1	0	0	?
60	1	1	1	?
50	0	0	?	0
30	0	1	?	1
40	0	0	?	0

- By randomization: fine to compare outcomes of treated with outcomes of untreated
- Based on baseline covariates (e.g., age): guesses about what outcome would be for all participants if they were (un)treated.
  - **By using the models that were used to obtain conditional estimates.**

A simple try...

- Let's use a simple imputation procedure:
  - Estimate disease risk on treatment, $\hat{P}^1$, for all trial participants based on a logistic regression in the treated, in function of baseline covariates.

Age	Z	Y	Y^1	$\hat{P}^1$	Y^0
40	1	1	1	0.8	?
50	1	0	0	0.7	?
60	1	1	1	0.6	?
50	0	0	?	0.7	0
30	0	1	?	0.9	1
40	0	0	?	0.8	0

  - average these risks for all trial participants to obtain an estimate of population disease risk on treatment (i.e., $E(Y^1)$).

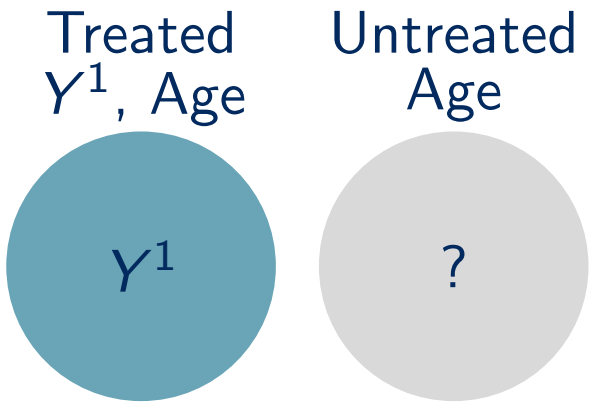
Covariate Adjusted Estimator: standardization/g-computation

Example: $E(Y^1)$

Treated	Untreated
Y^1, Age	Age

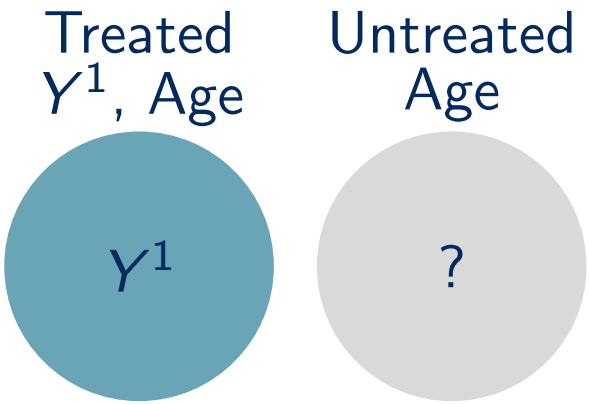
Covariate Adjusted Estimator: standardization/g-computation

Example: $E(Y^1)$



Covariate Adjusted Estimator: standardization/g-computation

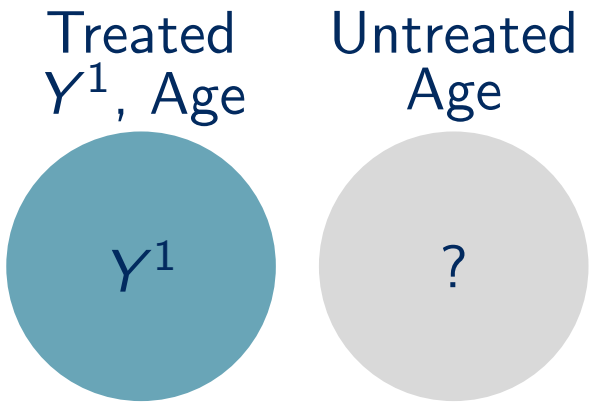
Example: $E(Y^1)$



Estimator for $E(Y^1)$ is obtained by

Covariate Adjusted Estimator: standardization/g-computation

Example: $E(Y^1)$



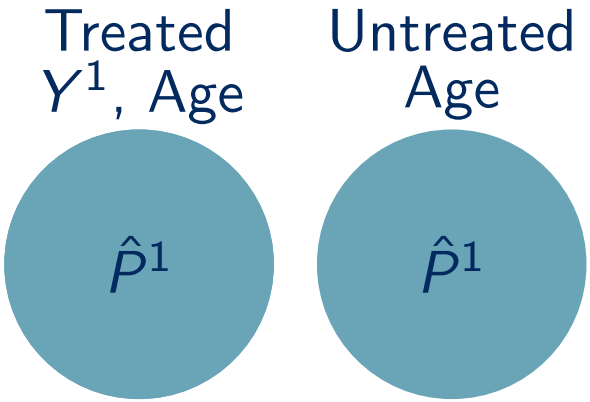
Estimator for $E(Y^1)$ is obtained by

Step 1: **Model fitting**

fitting a logistic regression model for outcome Y given age among the treated patients,

Covariate Adjusted Estimator: standardization/g-computation

Example: $E(Y^1)$



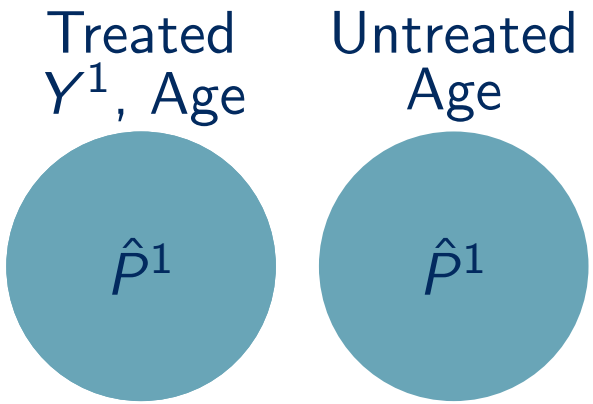
Estimator for $E(Y^1)$ is obtained by

Step 1: **Model fitting**
fitting a logistic regression model for outcome Y given age among the treated patients,

Step 2: **Predicting**
using this model to impute outcome for **all** patients,

Covariate Adjusted Estimator: standardization/g-computation

Example: $E(Y^1)$



Estimator for $E(Y^1)$ is obtained by

Step 1: **Model fitting**
fitting a logistic regression model for outcome Y given age among the treated patients,

Step 2: **Predicting**
using this model to impute outcome for **all** patients,

Step 3: **Averaging**
taking the average of imputed outcomes.

Covariate Adjusted Estimator: standardization/g-computation

- Similar for an estimate of population disease risk on control:
  - Estimate disease risk on control, $\hat{P}^0$, for all trial participants based on a logistic regression in the controls, in function of baseline covariates.

Age	Z	Y	Y^1	$\hat{P}^1$	Y^0	$\hat{P}^0$
40	1	1	1	0.8	?	0.3
50	1	0	0	0.7	?	0.2
60	1	1	1	0.6	?	0.1
50	0	0	?	0.7	0	0.2
30	0	1	?	0.9	1	0.4
40	0	0	?	0.8	0	0.3

  - average these risks for all trial participants to obtain an estimate of population disease risk on control (i.e., $E(Y^0)$).
- We can then contrast these estimates as differences, ratios, ...

Some Advantages

- Focus on marginal treatment effect leads to a simple interpretation
  - Same as comparing sample averages
  - No matter how complex logistic regression models are

Some Advantages

- Focus on marginal treatment effect leads to a simple interpretation

  Same as comparing sample averages

  No matter how complex logistic regression models are

- More efficient than standard sample averages if age is predictive for outcome
  - By contrasting disease risks for the same participants with and without treatment, we gain precision.

Simulation Results

Results for binary outcome and risk difference under correctly specified models

n	Effect	Estimator type	Bias	Power	MSE	RE
100	-0.201	Unadj.	0.025	0.463	0.829	1.000
		Adj.	0.023	0.607	0.755	0.911
200	-0.201	Unadj.	0.010	0.821	0.864	1.000
		Adj.	-0.001	0.895	0.749	0.867
500	-0.126	Unadj.	-0.013	0.798	0.979	1.000
		Adj.	-0.007	0.862	0.850	0.868
1000	-0.091	Unadj.	0.012	0.837	0.898	1.000
		Adj.	0.020	0.892	0.817	0.910

Results from Benkeser et al. (2020) "Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes." Biometrics.

Data Analysis: MISTIE II trial (Stroke)

- Participants were randomized to the treatment arm (surgical) or control arm (standard medical care).
- Randomization ratio was 2:1 treatment (66) to control (37).
- Functional outcome: proportion of patients who achieved a modified Rankin Scale score of 0-3 at 365 days (**binary**).
- Estimand of interest: **risk difference**.
- The following baseline variables are strongly associated with the primary outcome: age, ICH volume, and National Institutes of Health Stroke Scale (NIHSS).

(Hanley et al., 2016)

Data Analysis: MISTIE II trial

- **Unadjusted** estimator: difference between the observed proportion of successes in treatment versus control.
  - Estimate: 12.0%
  - 95% CI: -5.9% to 30.2%

Data Analysis: MISTIE II trial

- **Unadjusted** estimator: difference between the observed proportion of successes in treatment versus control.
  - Estimate: 12.0%
  - 95% CI: −5.9% to 30.2%
- **Covariate adjusted** estimator (TMLE)
  - Estimate: 14.4%
  - 95% CI: 1.3% to 32.8%

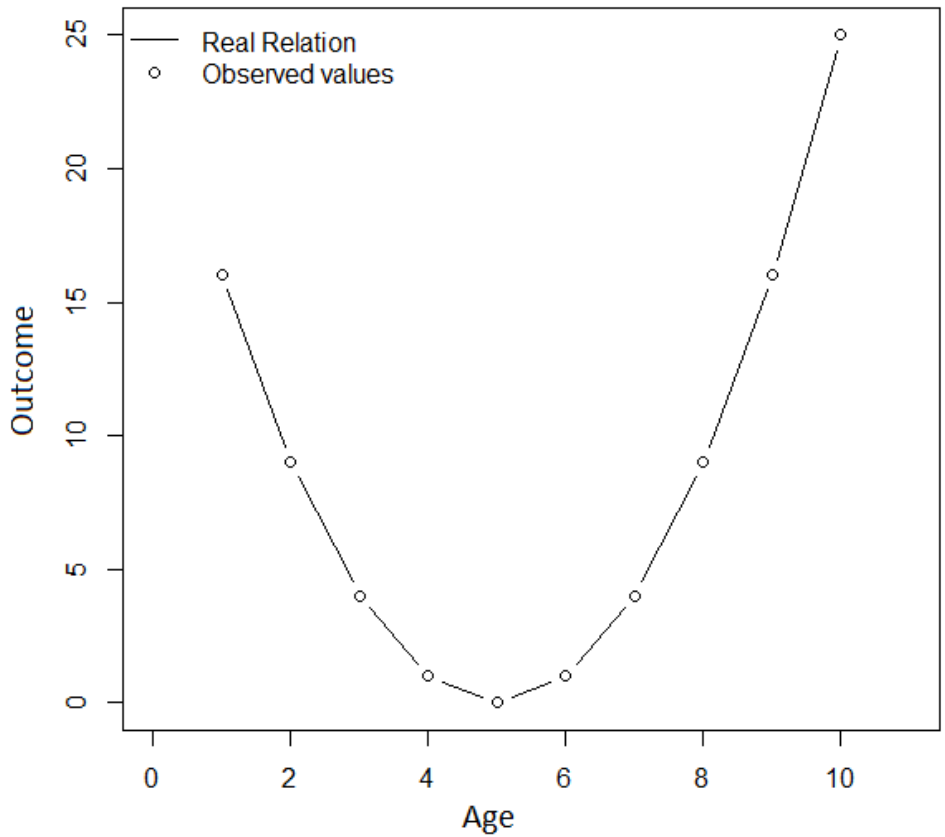
Data Analysis: MISTIE II trial

- **Unadjusted** estimator: difference between the observed proportion of successes in treatment versus control.
  - Estimate: 12.0%
  - 95% CI: −5.9% to 30.2%
- **Covariate adjusted** estimator (TMLE)
  - Estimate: 14.4%
  - 95% CI: 1.3% to 32.8%
- The width of this confidence interval is 12.7% **smaller** than that of the unadjusted estimator.

(Colantuoni and Rosenblum, 2015)

What if models are misspecified?

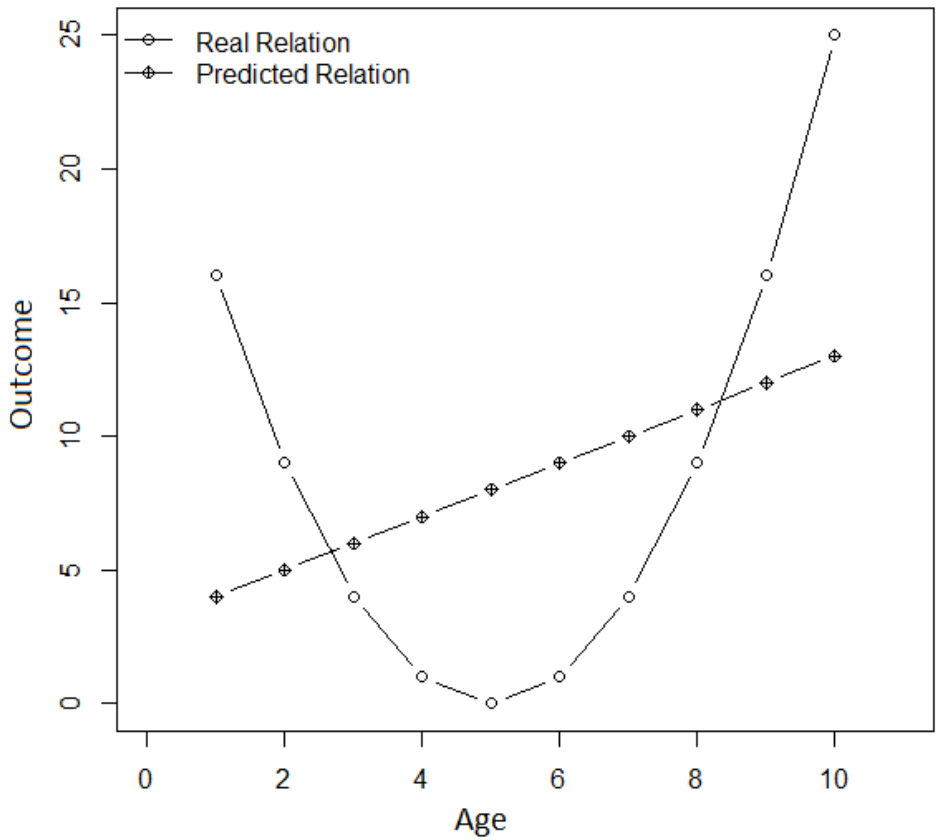
What if relationship between age and outcome in treated patients is not linear...



For simplicity, the outcome is continuous now

What if models are misspecified?

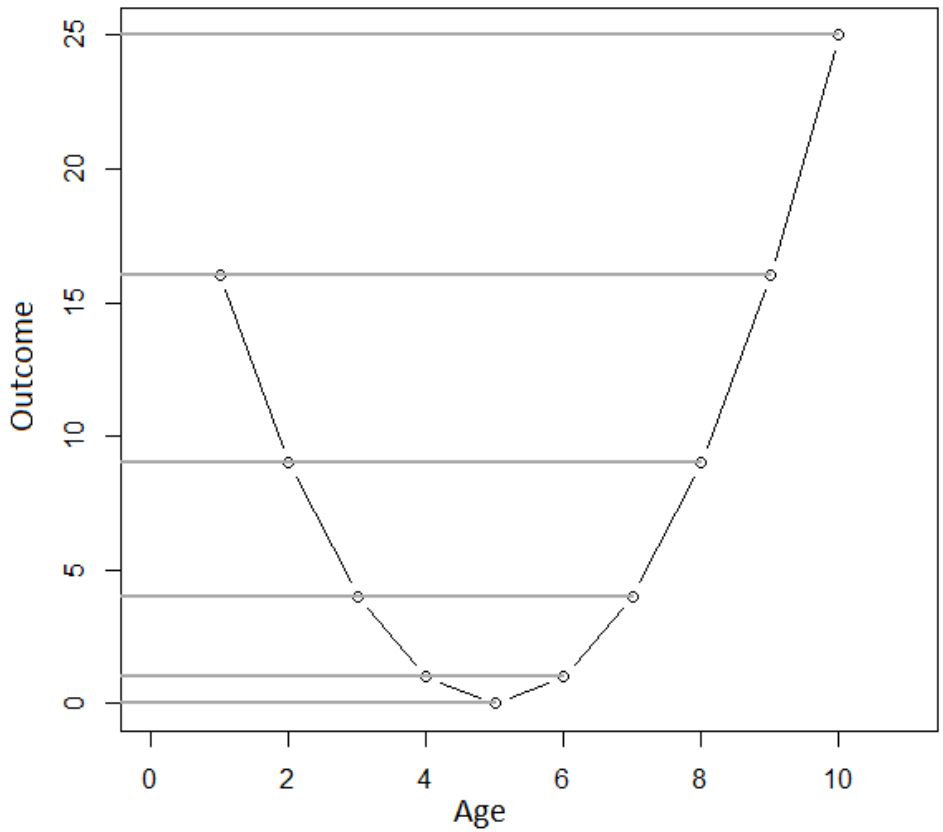
. . . , but we fit a misspecified model *outcome* $\sim$ *age*?



For simplicity, the outcome is continuous now

What if models are misspecified?

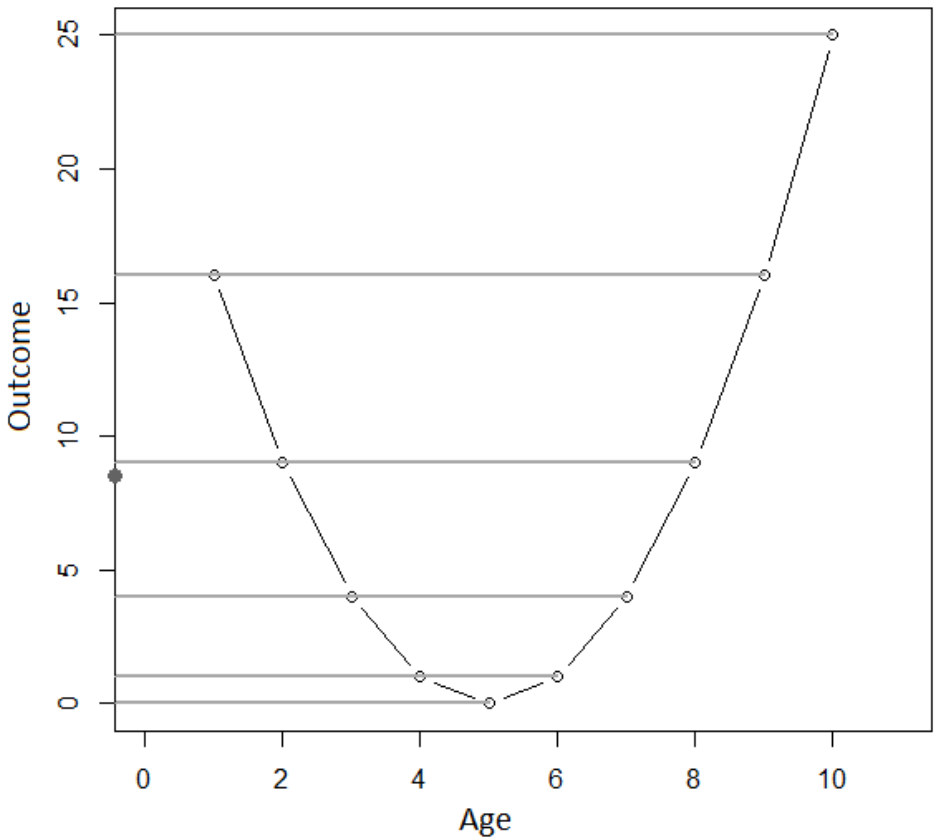
Projections of the observed outcomes on the y-axis,



For simplicity, the outcome is continuous now

What if models are misspecified?

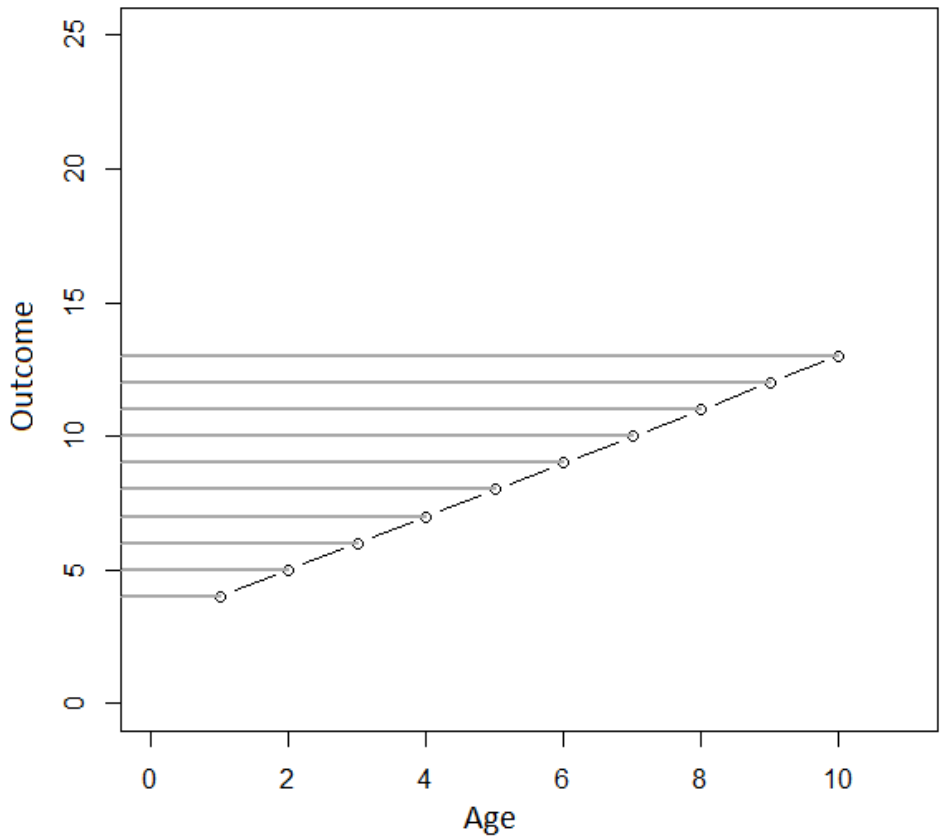
average to 8.5.



For simplicity, the outcome is continuous now

What if models are misspecified?

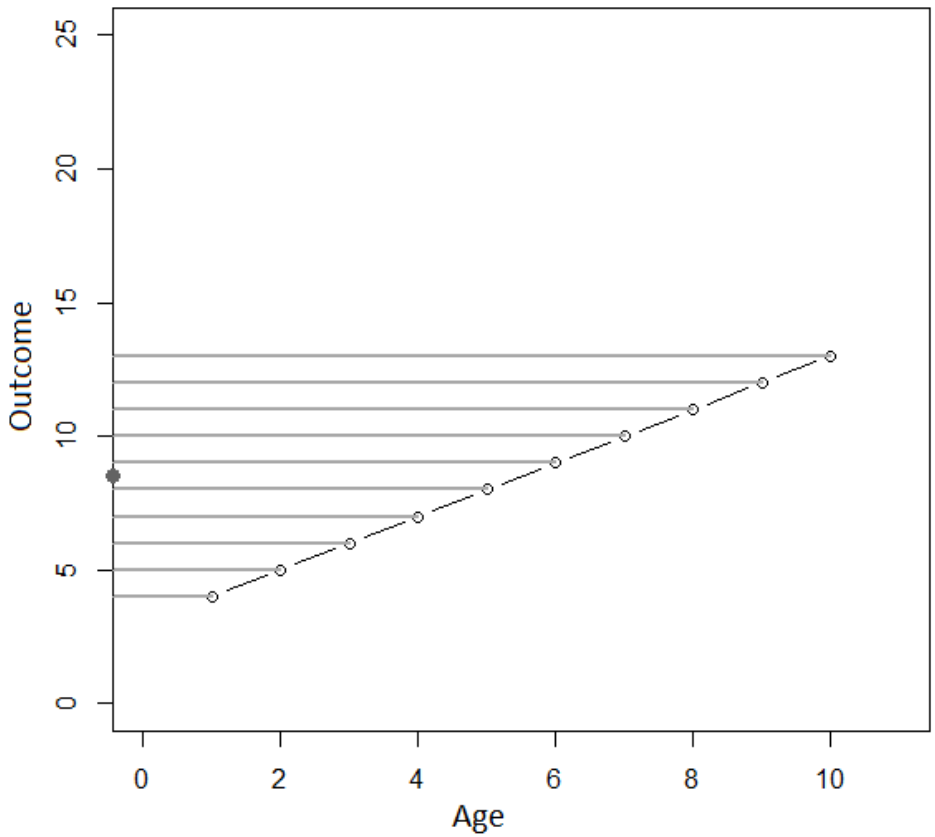
Projections of the predictions on the y-axis,



For simplicity, the outcome is continuous now

What if models are misspecified?

also average to 8.5.



For simplicity, the outcome is continuous now

What if models are misspecified?

- In treatment arm: mean of predictions (under treatment) = mean of observed outcomes,
  regardless of whether your model is correct or not

What if models are misspecified?

- In treatment arm: mean of predictions (under treatment) = mean of observed outcomes,
  regardless of whether your model is correct or not
- **Under randomization**, this robustness against misspecification also holds for mean of predictions (under treatment) for **all** patients

What if models are misspecified?

- In treatment arm: mean of predictions (under treatment) = mean of observed outcomes,
  regardless of whether your model is correct or not
- **Under randomization**, this robustness against misspecification also holds for mean of predictions (under treatment) for **all** patients

$\Rightarrow$ **Consistent estimator** for $E(Y^1)$, even when model is wrong.

Potential of baseline covariates

Mean of predictions based on glm's with canonical link and intercept, fitted in both arms separately

- Asymptotically unbiased estimator, even when outcome regression model is wrong (**robustness**)
  - They overcome the concern as to whether covariate adjustment (and possible misspecification of the model) is appropriate in randomized experiments.

Potential of baseline covariates

Mean of predictions based on glm's with canonical link and intercept, fitted in both arms separately

- Asymptotically unbiased estimator, even when outcome regression model is wrong (**robustness**)
  - They overcome the concern as to whether covariate adjustment (and possible misspecification of the model) is appropriate in randomized experiments.
- Model misspecification may reduce efficiency, but (almost) never outperformed by unadjusted analyses (**more efficient**).

Inference

- Standard errors easy to calculate

1. Robust standard errors (Tsiatis et al., 2008; Rosenblum and Van Der Laan, 2009; Ye et al., 2023):
   - Similar to variance of sample mean

     $1/n$ times sample variance of $2Z(Y - \hat{P}^1) + \hat{P}^1 - (2(1 - Z)(Y - \hat{P}^0) + \hat{P}^0)$ for a mean difference
   - Takes into account uncertainty in imputations

Inference

- Standard errors easy to calculate
  1. Robust standard errors (Tsiatis et al., 2008; Rosenblum and Van Der Laan, 2009; Ye et al., 2023):
     - Similar to variance of sample mean

       $1/n$ times sample variance of $2Z(Y - \hat{P}^1) + \hat{P}^1 - (2(1 - Z)(Y - \hat{P}^0) + \hat{P}^0)$ for a mean difference
     - Takes into account uncertainty in imputations
  2. Non-parametric bootstrap
- and are valid even when the model is misspecified (Vermeulen et al., 2015)

Inference

- Standard errors easy to calculate
  1. Robust standard errors (Tsiatis et al., 2008; Rosenblum and Van Der Laan, 2009; Ye et al., 2023):
     - Similar to variance of sample mean

       $1/n$ times sample variance of $2Z(Y - \hat{P}^1) + \hat{P}^1 - (2(1 - Z)(Y - \hat{P}^0) + \hat{P}^0)$ for a mean difference
     - Takes into account uncertainty in imputations
  2. Non-parametric bootstrap
- and are valid even when the model is misspecified (Vermeulen et al., 2015)
- Robust standard errors also valid when variable selection is used (Avagyan and Vansteelandt, 2021).

Recommendations

- Important to use predictions based on **glm's with canonical link**.
  - Otherwise we need slightly **different approach** (AIPW, TMLE).

Recommendations

- Important to use predictions based on **glm's with canonical link**.
  - Otherwise we need slightly **different approach** (AIPW, TMLE).
- Use of baseline covariates raises concerns due to **missing data**
  - Easily addressed: **mean/mode imputation**.
  - Without inflating risk of bias.

Recommendations

- Important to use predictions based on **glm's with canonical link**.
  - Otherwise we need slightly **different approach** (AIPW, TMLE).
- Use of baseline covariates raises concerns due to **missing data**
  - Easily addressed: **mean/mode imputation**.
  - Without inflating risk of bias.
- I haven't covered all available methods
  - There are no other methods that have more power and have the same robustness.

What about hypothesis testing (p-value)?

- Suppose we are fitting a generalised linear model with pre-specified canonical link function $g(\cdot)$

$$g\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X,$$

using maximum likelihood estimation.

What about hypothesis testing (p-value)?

- Suppose we are fitting a generalised linear model with pre-specified canonical link function $g(\cdot)$

$$g\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X,$$

  using maximum likelihood estimation.

- Wald tests (using robust SE) based on
  1. $\hat{\beta}_1$ (conditional), or
  2. standardization with this model (marginal),

  both **control the Type I error** rate and are **equally powerful** in large samples (Rosenblum and Steingrimsson, 2016).

What about hypothesis testing (p-value)?

- Suppose we are fitting a generalised linear model with pre-specified canonical link function $g(\cdot)$

$$g\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X,$$

  using maximum likelihood estimation.

- Wald tests (using robust SE) based on
  1. $\hat{\beta}_1$ (conditional), or
  2. standardization with this model (marginal),

  both **control the Type I error** rate and are **equally powerful** in large samples (Rosenblum and Steingrimsson, 2016).

- Enhanced standardization estimators (e.g., by fitting separate outcome working models or by including a model for randomization) have the potential for greater **efficiency gains**.

Thank you for your attention!

Interested? ▸ Paper with Frank Bretz and Oliver Dukes and ▸ Tutorials

E-mail: kelly.vanlancker@ugent.be
Website: kellyvanlancker.com

References I

Avagyan, V. and S. Vansteelandt (2021). High-dimensional inference for the average treatment effect under model misspecification using penalized bias-reduced double-robust estimation. *Biostatistics & Epidemiology*, 1–18.

Benkeser, D., I. Díaz, A. Luedtke, J. Segal, D. Scharfstein, and M. Rosenblum (2020). Improving precision and power in randomized trials for covid-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*.

Colantuoni, E. and M. Rosenblum (2015). Leveraging prognostic baseline variables to gain precision in randomized trials. *Statistics in medicine 34*(18), 2602–2617.

References II

Daniel, R., J. Zhang, and D. Farewell (2021). Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal 63*(3), 528–557.

EMA (2015). Guideline on adjustment for baseline covariates in clinical trials. Last checked: 2022-05-30.

FDA (2023). Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products. Guidance for Industry. https://www.fda.gov/media/148910/download. Last checked: 2021-10-20.

References III

FDA and EMA (1998). E9 statistical principles for clinical trials. U.S. Food and Drug Administration: CDER/CBER. European Medicines Agency: CPMP/ICH/363/96. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf. Last checked: 2021-02-03.

Hanley, D. F., R. E. Thompson, J. Muschelli, M. Rosenblum, N. McBee, K. Lane, A. J. Bistran-Hall, S. W. Mayo, P. Keyl, D. Gandhi, et al. (2016). Safety and efficacy of minimally invasive surgery plus recombinant tissue plasminogen activator in intracerebral haemorrhage evacuation (mistie): a randomised, phase 2 trial. *The Lancet. Neurology 15*(12), 1228.

References IV

Jiang, F., L. Tian, H. Fu, T. Hasegawa, and L. J. Wei (2018). Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study. *Journal of the American Statistical Association 0*, 1–37.

Koch, G. G., C. M. Tangen, J.-W. Jung, and I. A. Amara (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Stat. Med. 17*(15-16), 1863–1892.

Moore, K. and M. J. van der Laan (2009a). Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. *Stat. Med. 28*(1), 39–64.

References V

Moore, K. L. and M. J. van der Laan (2009b). Increasing power in randomized trials with right censored outcomes through covariate adjustment. *Journal of Biopharmaceutical Statistics 19*(6), 1099–1131. PMID: 20183467.

Rosenblum, M. and J. A. Steingrimsson (2016). Matching the efficiency gains of the logistic regression estimator while avoiding its interpretability problems, in randomized trials.

Rosenblum, M. and M. J. Van Der Laan (2009). Using regression models to analyze randomized trials: Asymptotically valid hypothesis tests despite incorrectly specified models. *Biometrics 65*(3), 937–945.

References VI

Rubin, D. and M. van der Laan (2008). Covariate adjustment for the intention-to-treat parameter with empirical efficiency maximization. *U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 229*, https://biostats.bepress.com/ucbbiostat/paper229.

Tsiatis, A. A., M. Davidian, M. Zhang, and X. Lu (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in medicine 27*(23), 4658–4677.

Vermeulen, K., O. Thas, and S. Vansteelandt (2015). Increasing the power of the mann-whitney test in randomized experiments through flexible covariate adjustment. *Statistics in medicine 34*(6), 1012–1030.

References VII

Yang, L. and A. Tsiatis (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *The American Statistician 55*(4), 314–321.

Ye, T., M. Bannick, Y. Yi, and J. Shao (2023). Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. *Statistical Theory and Related Fields*, 1–5.

Ye, T., J. Shao, Y. Yi, and Q. Zhao (2022). Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Journal of the American Statistical Association*, 1–13.

Zhang, M. (2015, Jan). Robust methods to improve efficiency and reduce bias in estimating survival curves in randomized clinical trials. *Lifetime Data Analysis 21*(1), 119–137.

Marginal and conditional estimands

Arguments made for marginal estimands

- A single number with a (relatively) simple interpretation.
  - Yes, but we should not use that as an argument for a unadjusted analysis.
- Useful for making blanket policy decisions (e.g., should this drug be approved?)
  - Yes, but only if target population is similar to trial population.
- Less risk that model misspecification invalidates the analysis.
  - Used in defense of unadjusted analysis, or adjusted analysis for marginal estimands.

Marginal and conditional estimands

Arguments made for conditional estimands

- A broader understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.
  - Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity). This is not the case if no interactions are included.
  - Conditional estimands are more relevant to an individual.
- Estimators of conditional estimands are more precise.
  - This is an argument for adjusted analyses, rather than for conditional estimands.
- It is often argued that conditional estimands are more transportable to different populations.