

Statistically Speaking Lecture Series

Sponsored by the Biostatistics Collaboration Center

Considerations for Choosing Covariates in Clinical Research

Lauren C Balmert, PhD

Assistant Professor

Department of Preventive Medicine

December 9, 2021

Via Zoom

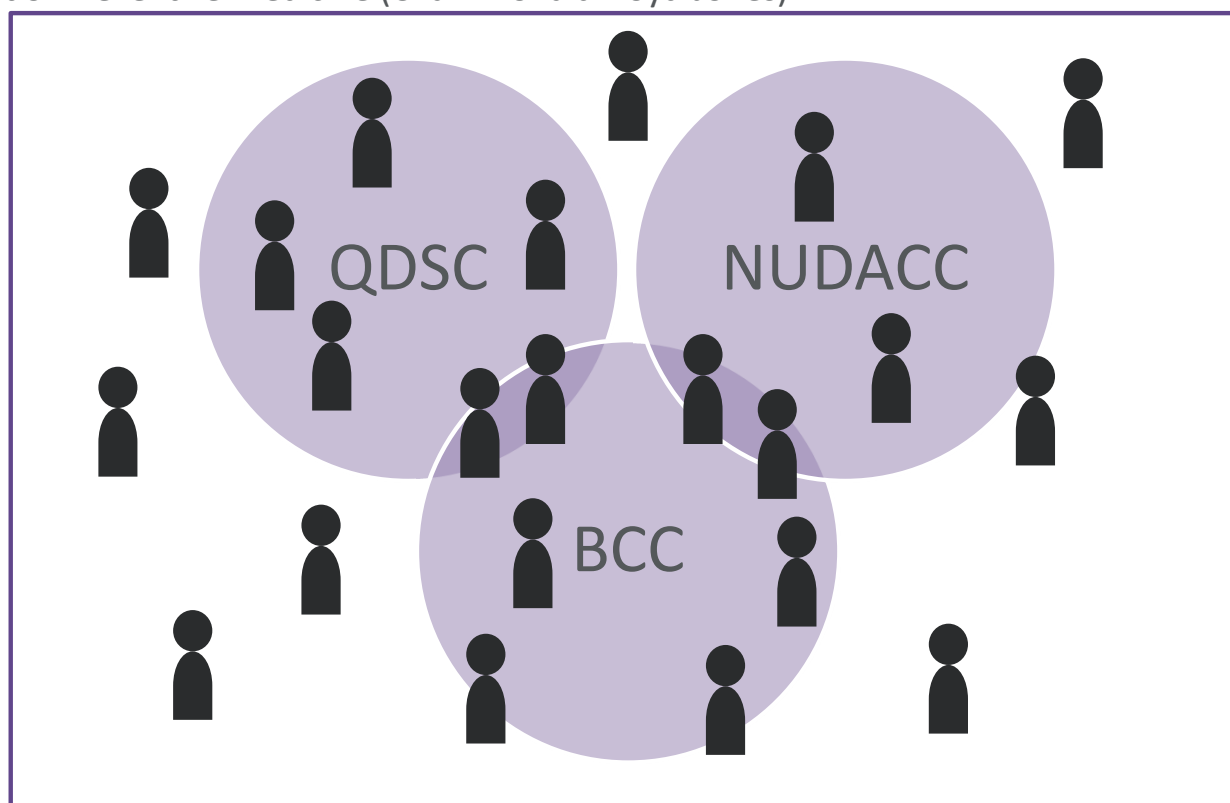
 **Northwestern Medicine**[®]
Feinberg School of Medicine



Biostatistics at NU

Overview

Division of Biostatistics (Chief: Denise Scholtens),
Department of Preventive Medicine (Chair: Donald Lloyd-Jones)



Biostatistics Centers and Cores

Overview



Biostatistics Collaboration Center (BCC)

- Supports **non-cancer** research at NU
- Initial 1-2 hour consultation subsidized by FSM Research Office
- Grant, Hourly
- <https://www.feinberg.northwestern.edu/sites/bcc/>

Quantitative Data Sciences Core (QDSC)

- Supports **cancer-related** research at NU
- Free to Lurie Cancer Center (LCC) members
- Grant
- <https://www.cancer.northwestern.edu/research/shared-resources/quantitative-data-sciences.html>

Northwestern University Data Analysis and Coordinating Center (NUDACC)

- Prospective, large **multicenter research**
- Comprehensive support (e.g., clinical monitoring, data analysis, project management)
- Grant
- <https://www.feinberg.northwestern.edu/sites/nudacc/>



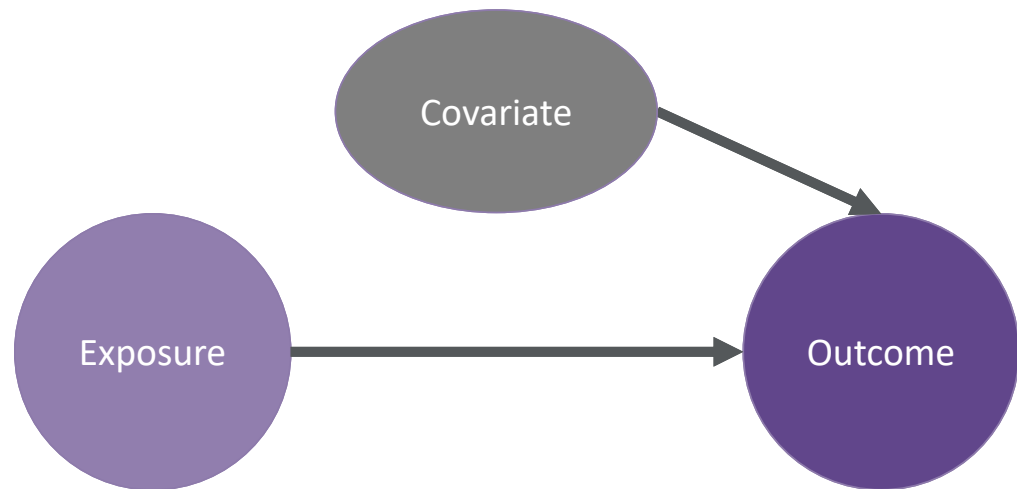
Considerations for Choosing Covariates in Clinical Research

Outline

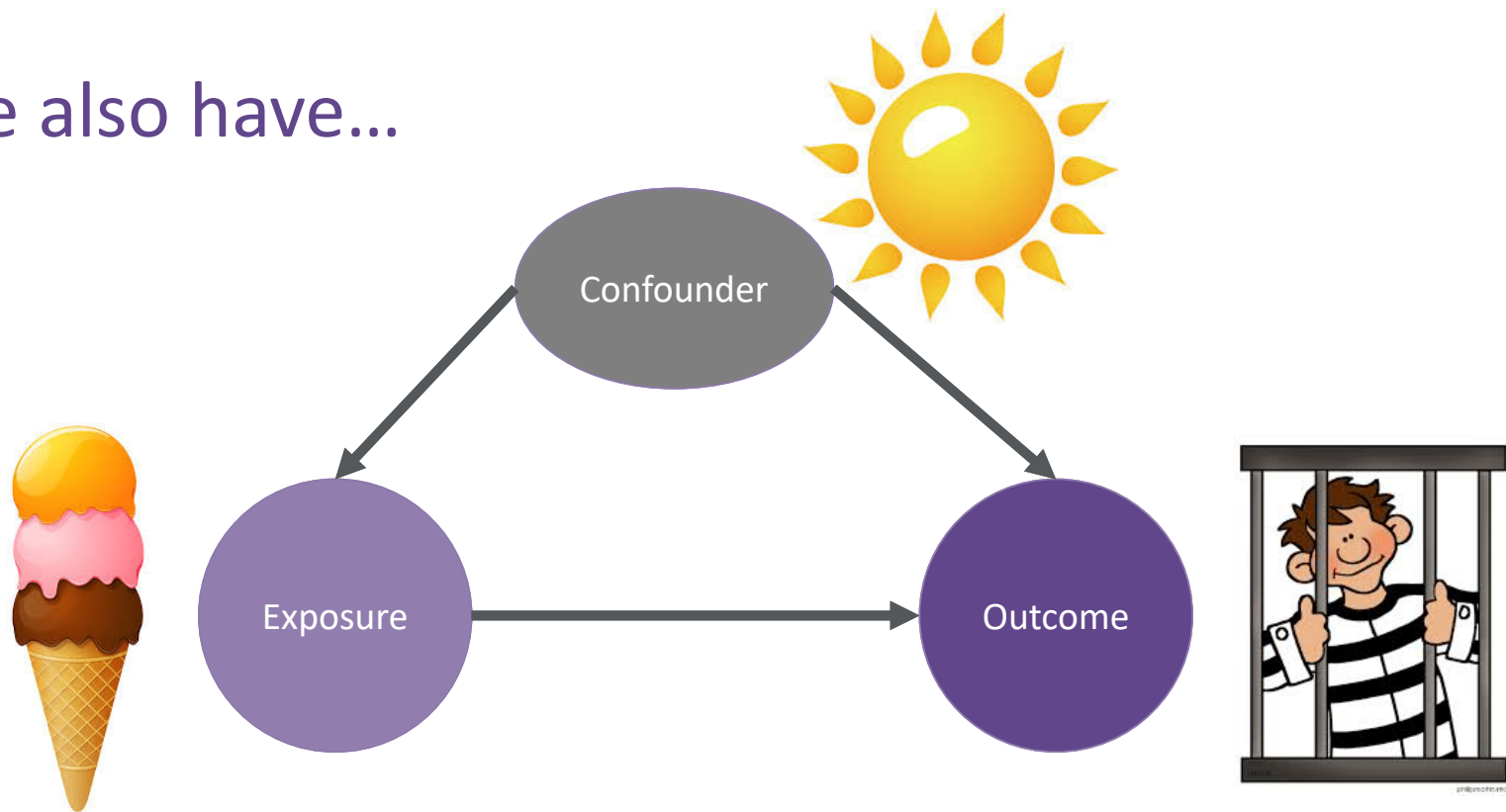
- General Concepts
- Observational studies
- Randomized clinical trials
- General Guidance

What is a covariate?

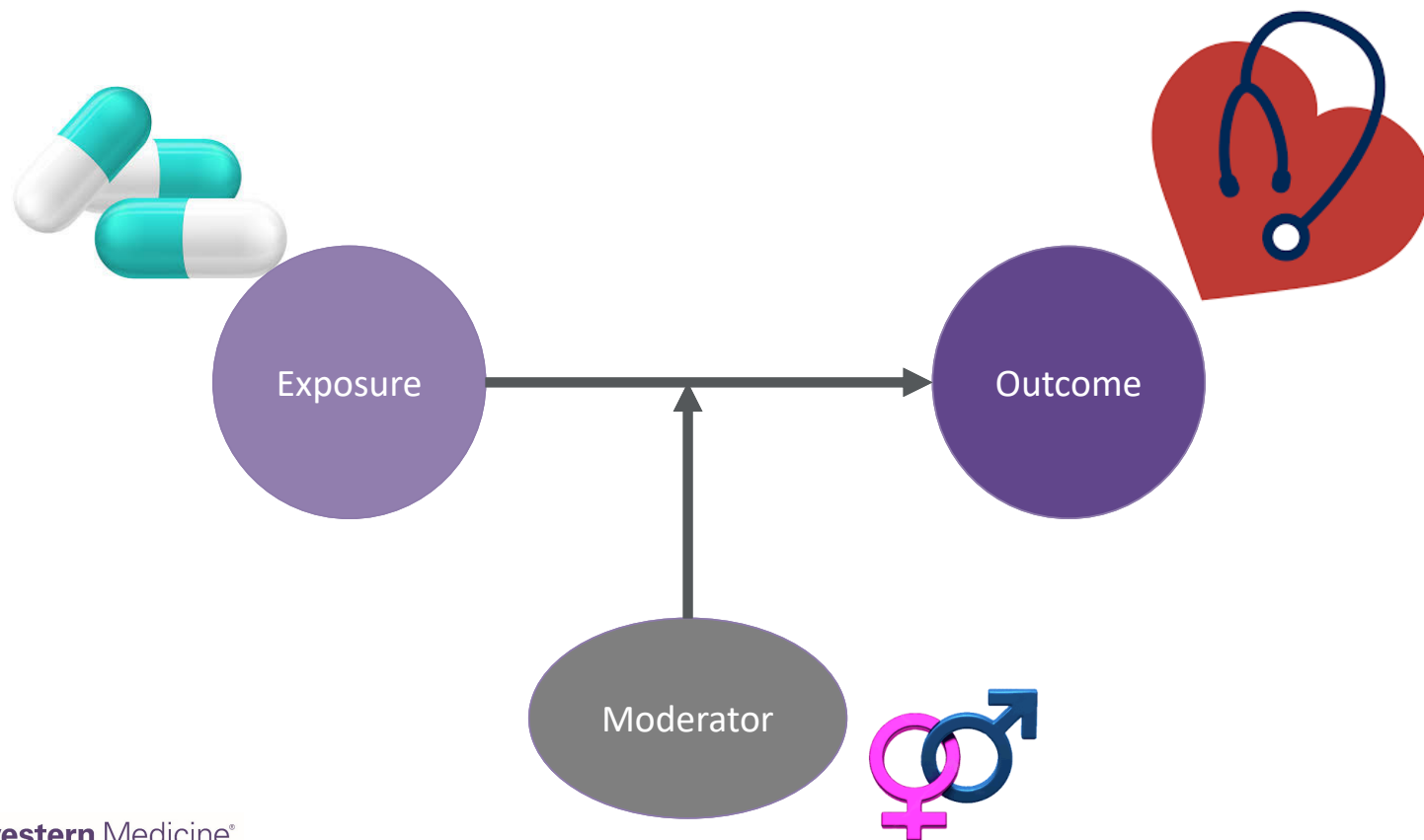
- Any variable that may be associated with an outcome (dependent variable)
 - Explain part of the variability in the outcome
- Not the outcome
- Not the exposure/intervention



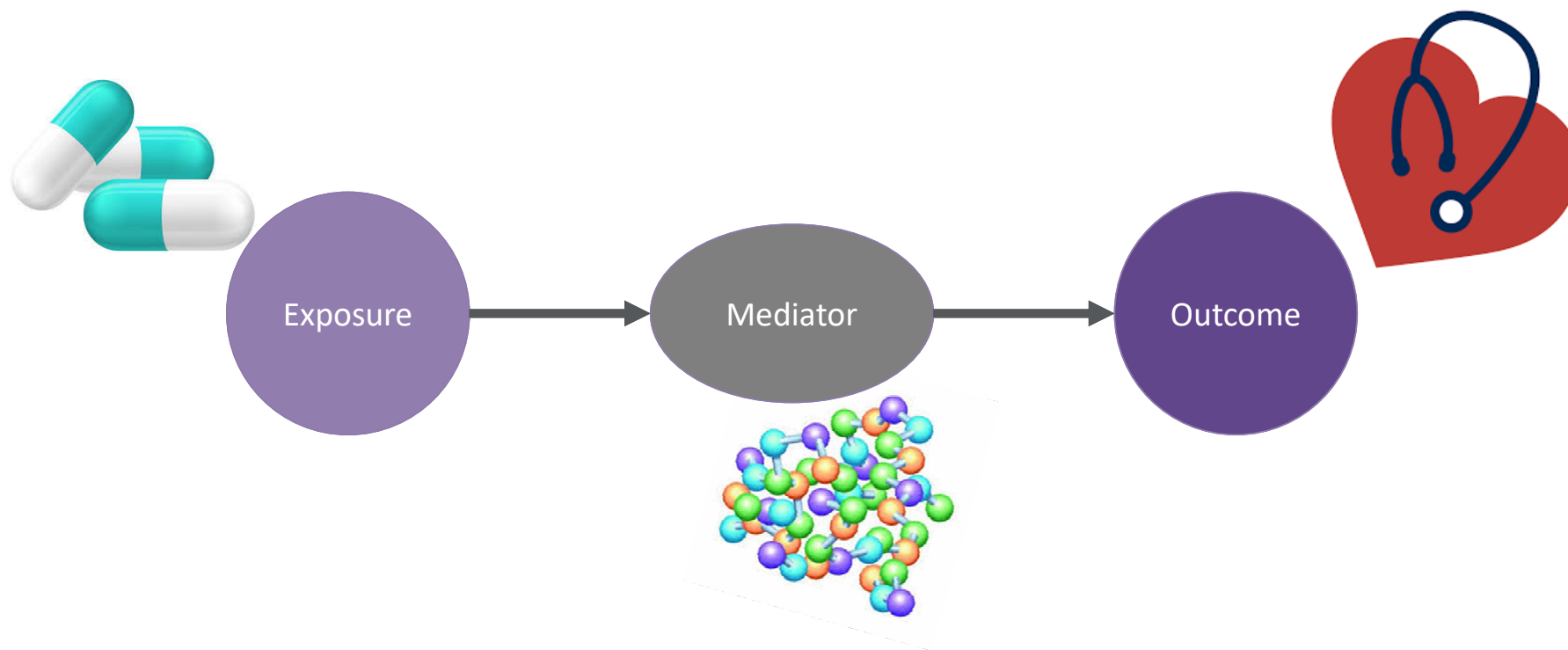
We also have...

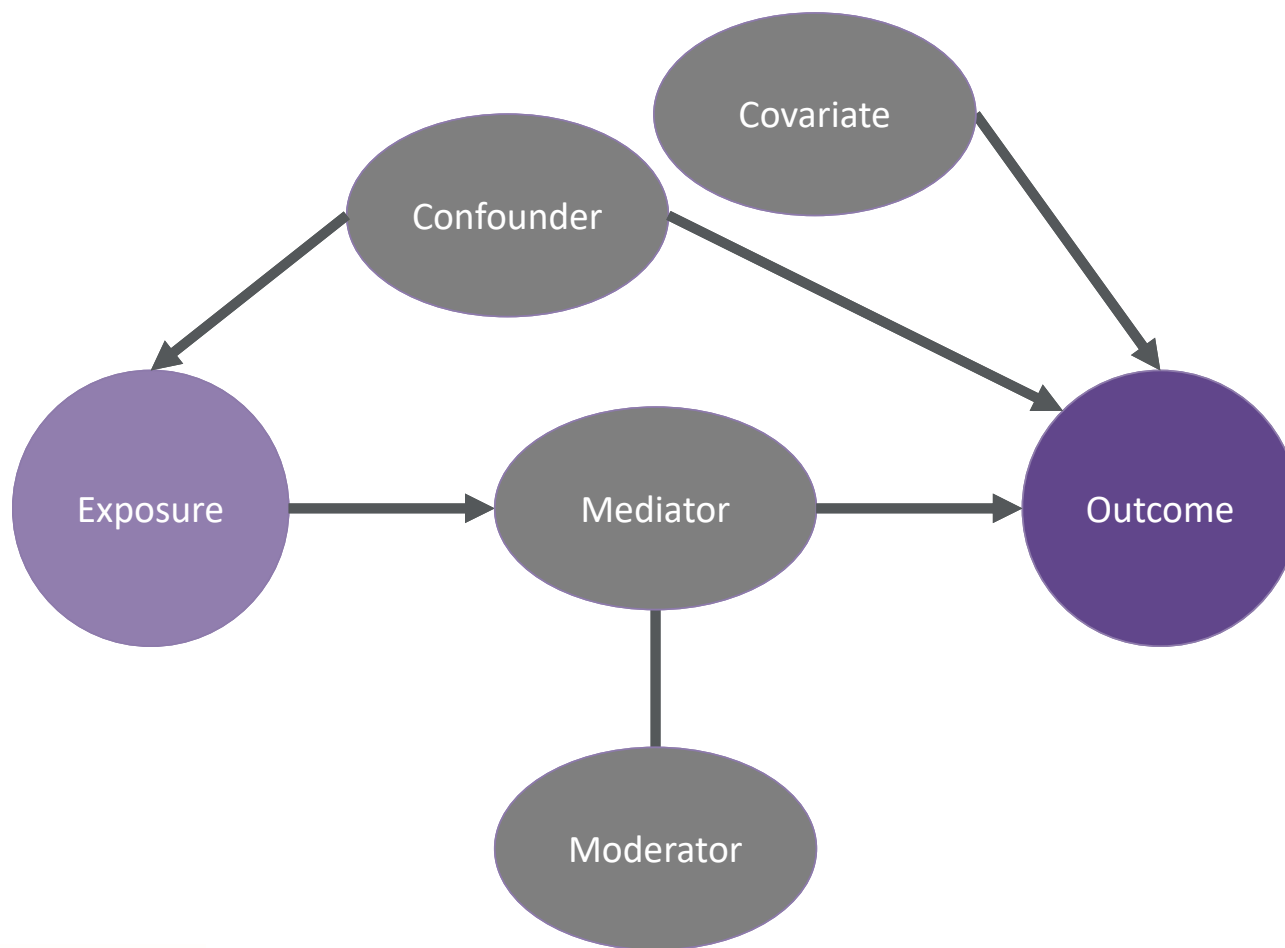


We also have...



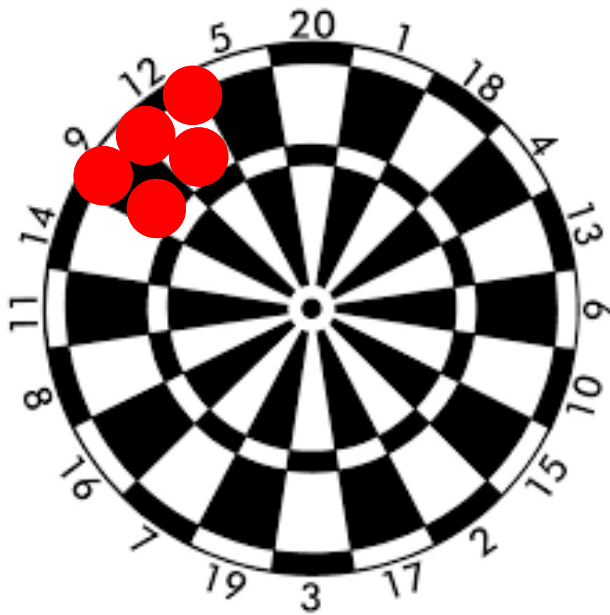
We also have...



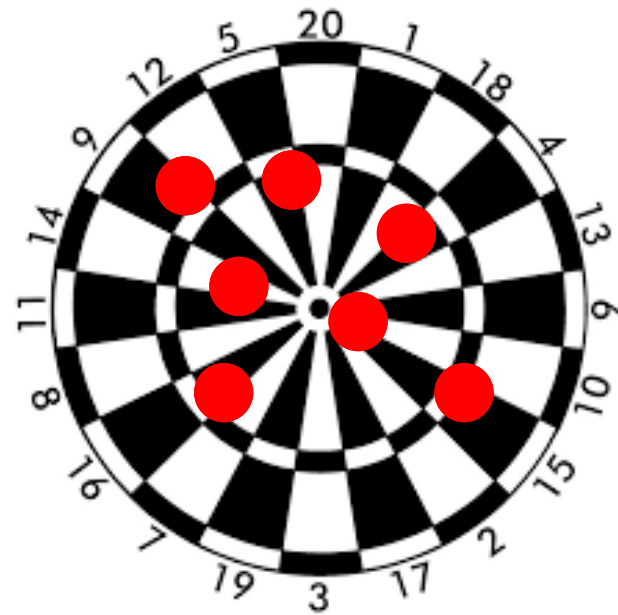


Why do I need to consider covariates?

Bias

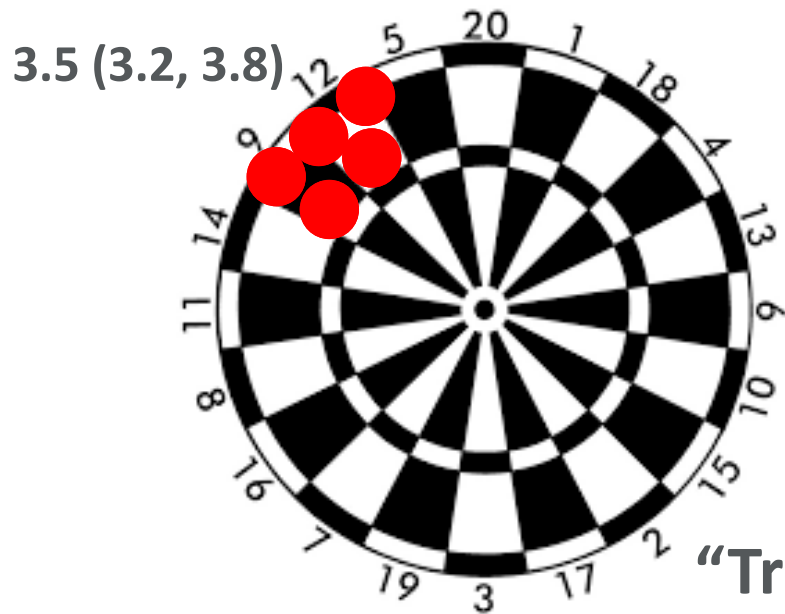


Precision

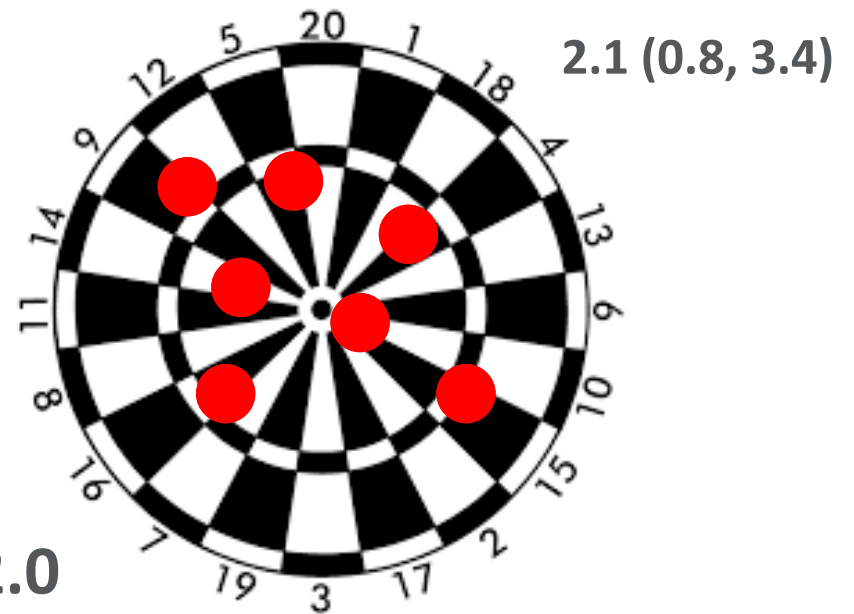


Why do I need to consider covariates?

Bias



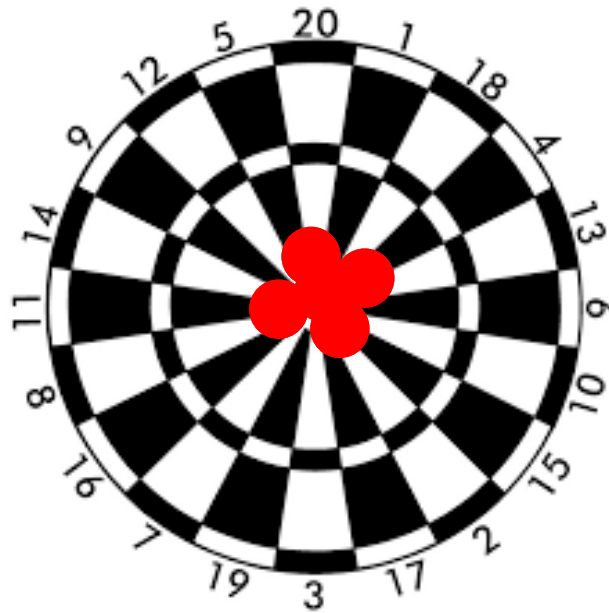
Precision



“True” OR=2.0

Ideally...

Low Bias, High Precision



Selection Bias

Information Bias

Recall Bias

Attrition Bias

Confounding Bias

Measurement Error

Sample Size

M Northwestern Medicine®
Feinberg School of Medicine

Considerations in Observational Studies



Explanatory vs. Prediction

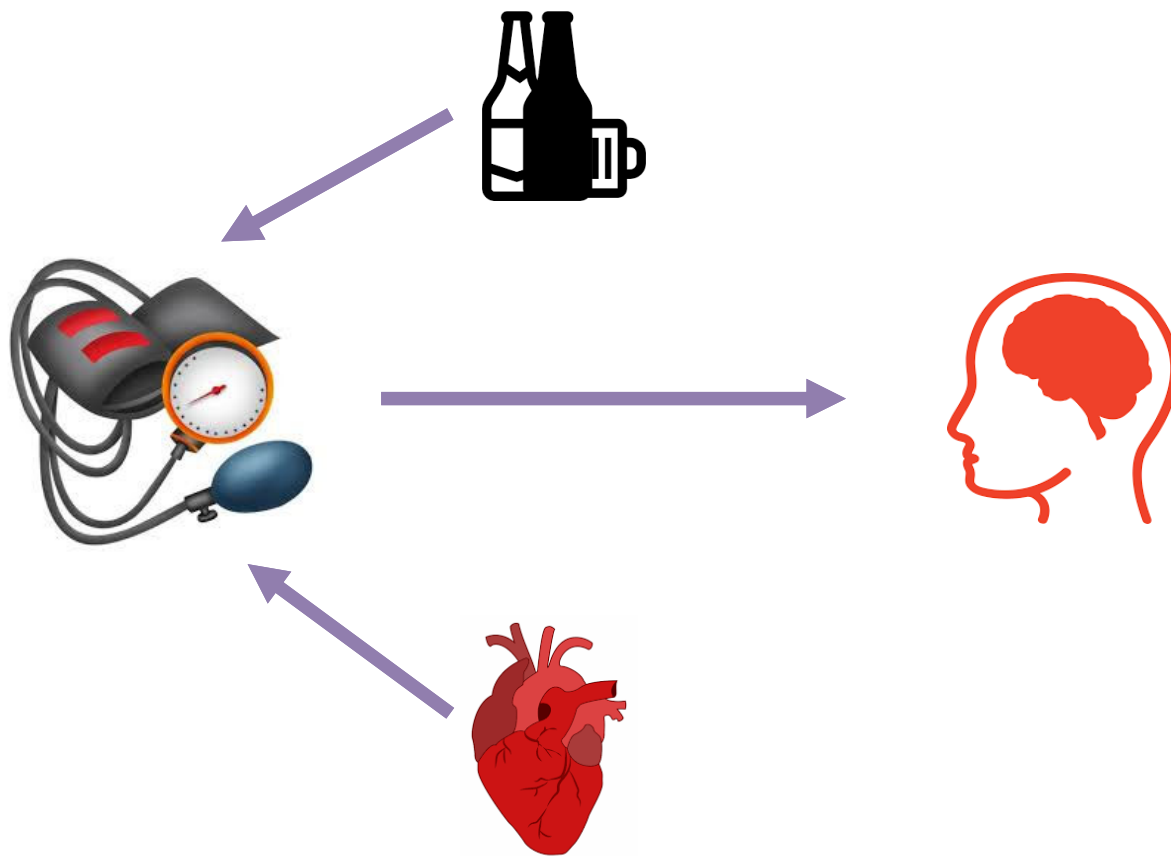
- Explanatory
 - Understanding the effects of an exposure / predictor
 - Interest in exposure effect estimate and significance
- Prediction
 - Accurate predictions of new/future observation

Why should I consider covariates?

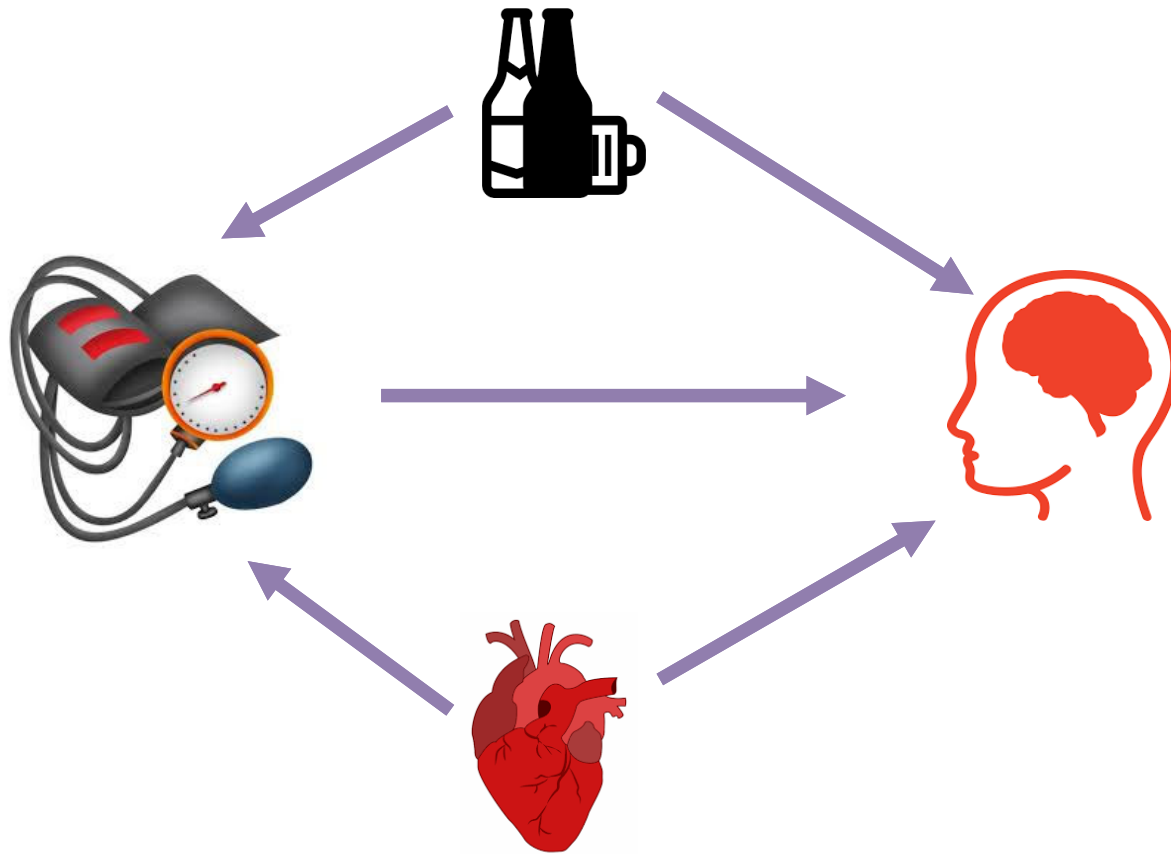
Is hypertension associated with an increased odds of stroke?



Why should I consider covariates?



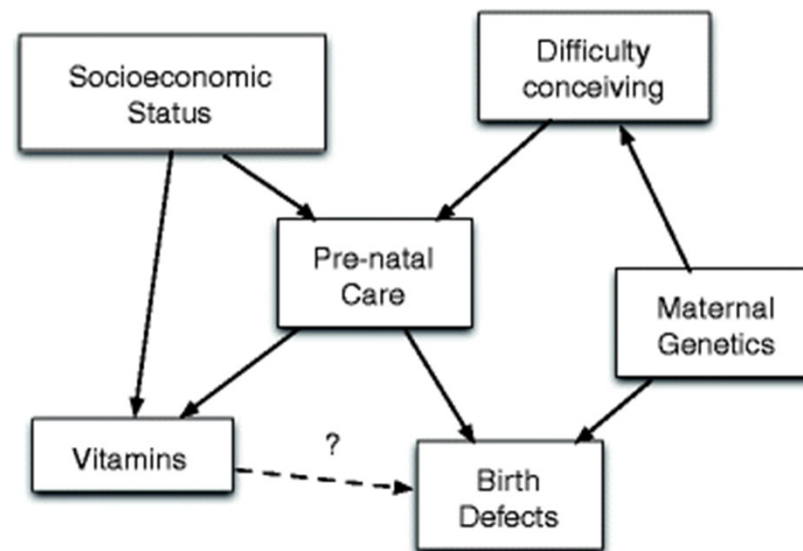
Why should I consider covariates?



How do we know what covariates to include?

- A priori selection
 - Known or hypothesized to be associated with the exposure and outcome
 - Draw a diagram to discern confounders, mediators, moderators, etc.

Directed Acyclic Graph
(DAG) for effects of
vitamin use on birth
defects



Vittinghoff et. al., Predictor Selection (Fig 10.2)

How do we know what covariates to include?

- A priori selection
 - Quality
 - Measurement error
 - Objective vs. subjective
 - Validated and reliable
 - Exclude those that are..
 - On the causal pathway (mediators)
 - Alternative measures of the exposure or outcome

How do we know what covariates to include?

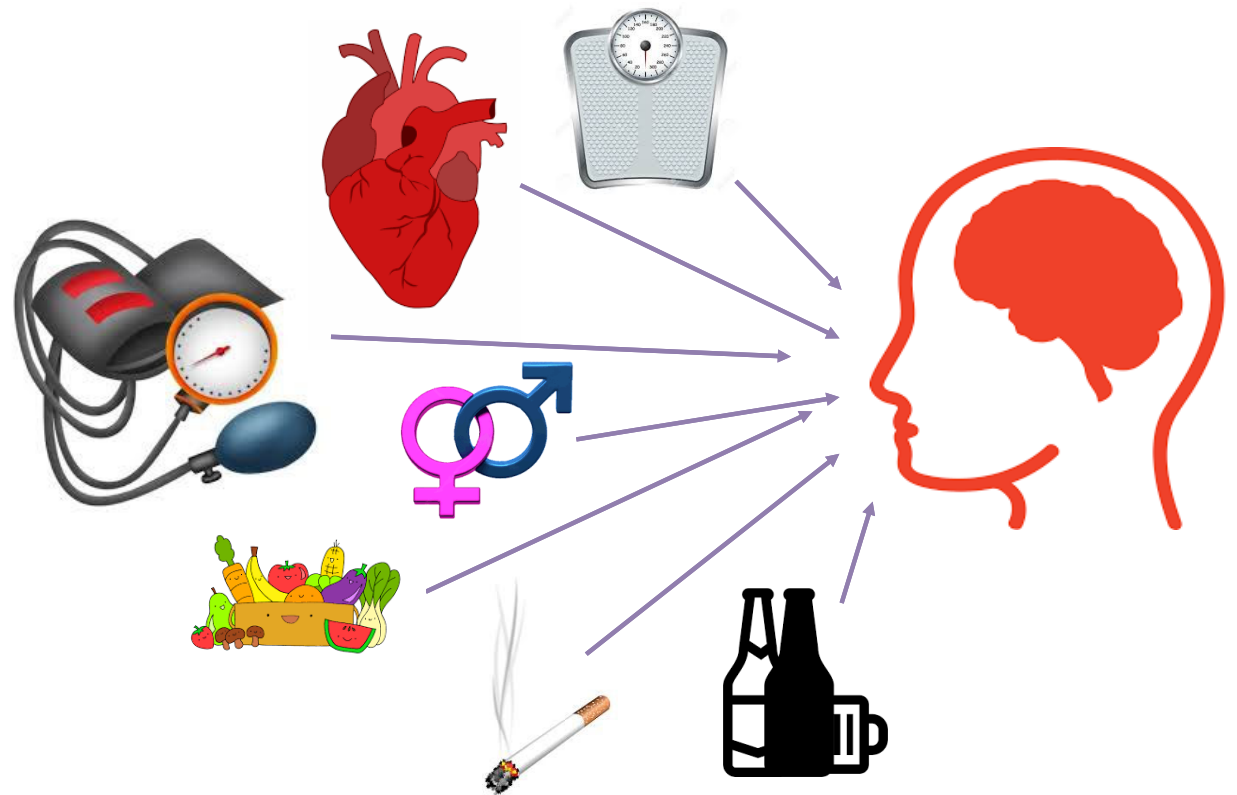
- Sample assessment
 - Missing data
 - Variability
- Data driven approaches
 - “Traditional” selection
 - Bivariate screening
 - Purposeful selection
 - Backwards / Forward / Stepwise
 - Cross-validation with subsets of candidate predictors
 - Penalized regression approaches

Concerns with data driven approaches

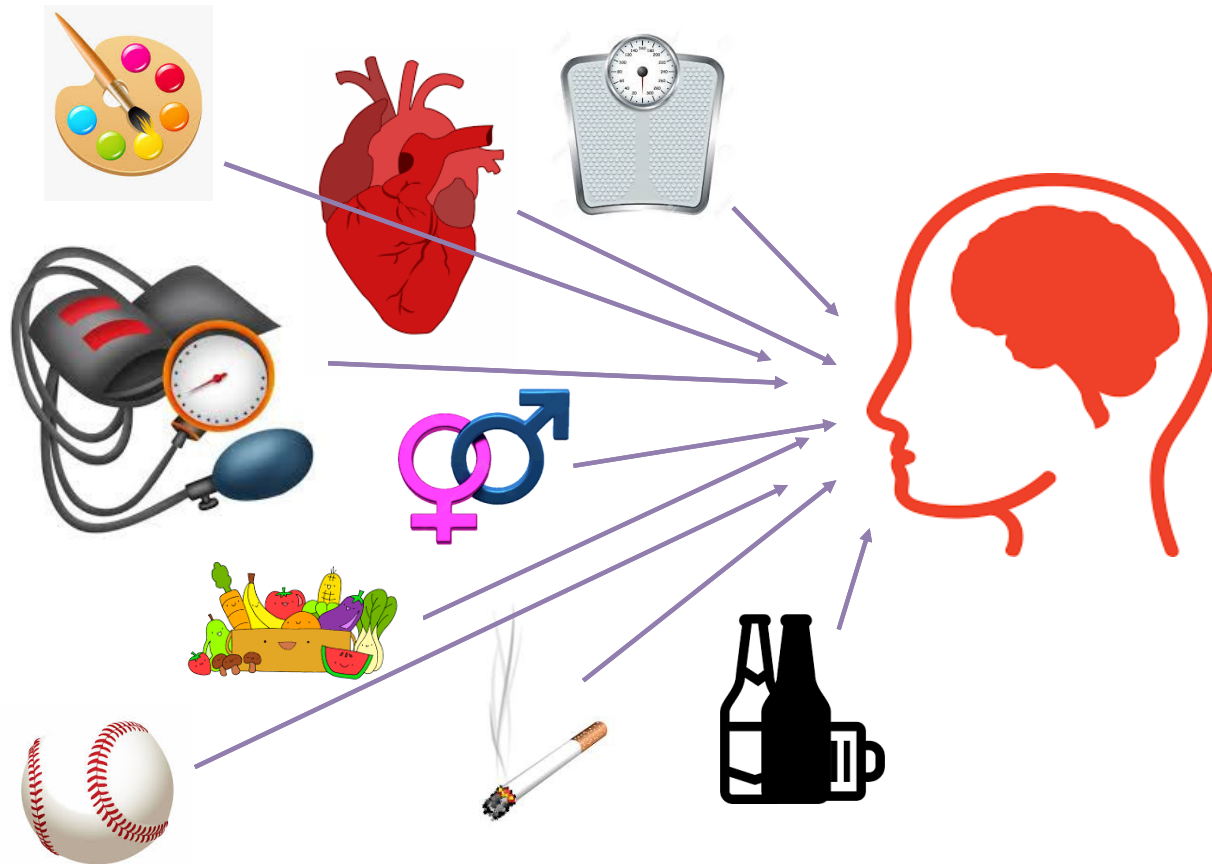
- Exclude potentially **important** confounders (small data sets)
- Include potentially **unimportant** confounders (large data sets)
- Exclude variables that are only **important after adjustment**
- Depends on **estimated effects** rather than true values
- Automated procedures **ignore clinical implications**
- Data have **already been used** to carry out covariate selection
- Increased probability of rejecting the null when it's true (**Type I error inflation**)

Why should I consider covariates?

Develop a model
to identify
adults at high
risk of stroke



Why should I consider covariates?



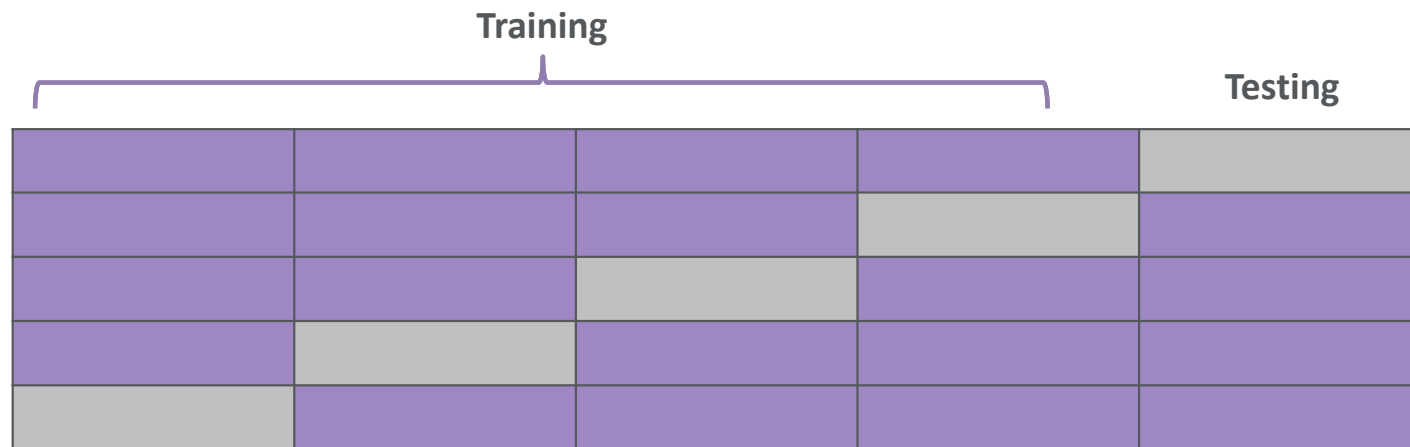
How do we know what covariates to include?

- Prespecifying covariates
- Practicality
 - Exclusion of invasive / risky / unreliable tests
 - Timing of assessment
- Sample assessment
 - Missing data
 - Variability



How do we know what covariates to include?

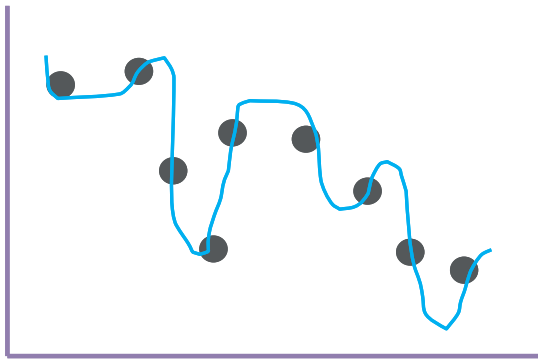
- Data-driven approaches
 - Cross-validation with subsets of candidate predictors
 - Penalized regression approaches
 - Random forests



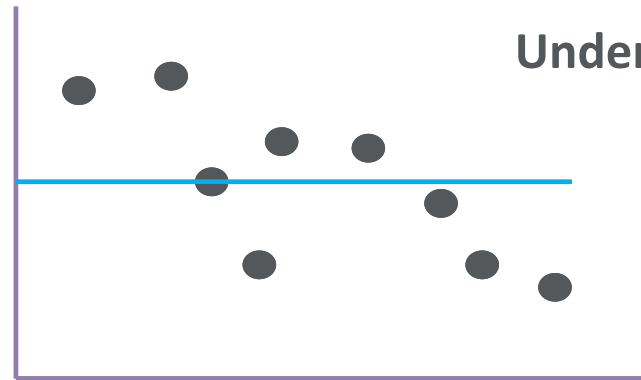
How many is too many?

- “Rule of Thumb”
 - 10 observations/outcomes per covariate
 - Caution with using this as a hard rule
- Prediction modeling approaches can generally incorporate more covariates
 - Want avoid overfitting

Overfitting



Underfitting



How many is too many?

- “Rule of Thumb”
 - 10 observations/outcomes per covariate
 - Caution with using this as a hard rule
- Prediction modeling approaches can generally incorporate more covariates
 - Want avoid overfitting
- High dimensional data
 - Penalized regression approaches (Predictors > Sample Size)
 - Variable reduction methods (PCA)

General guidance

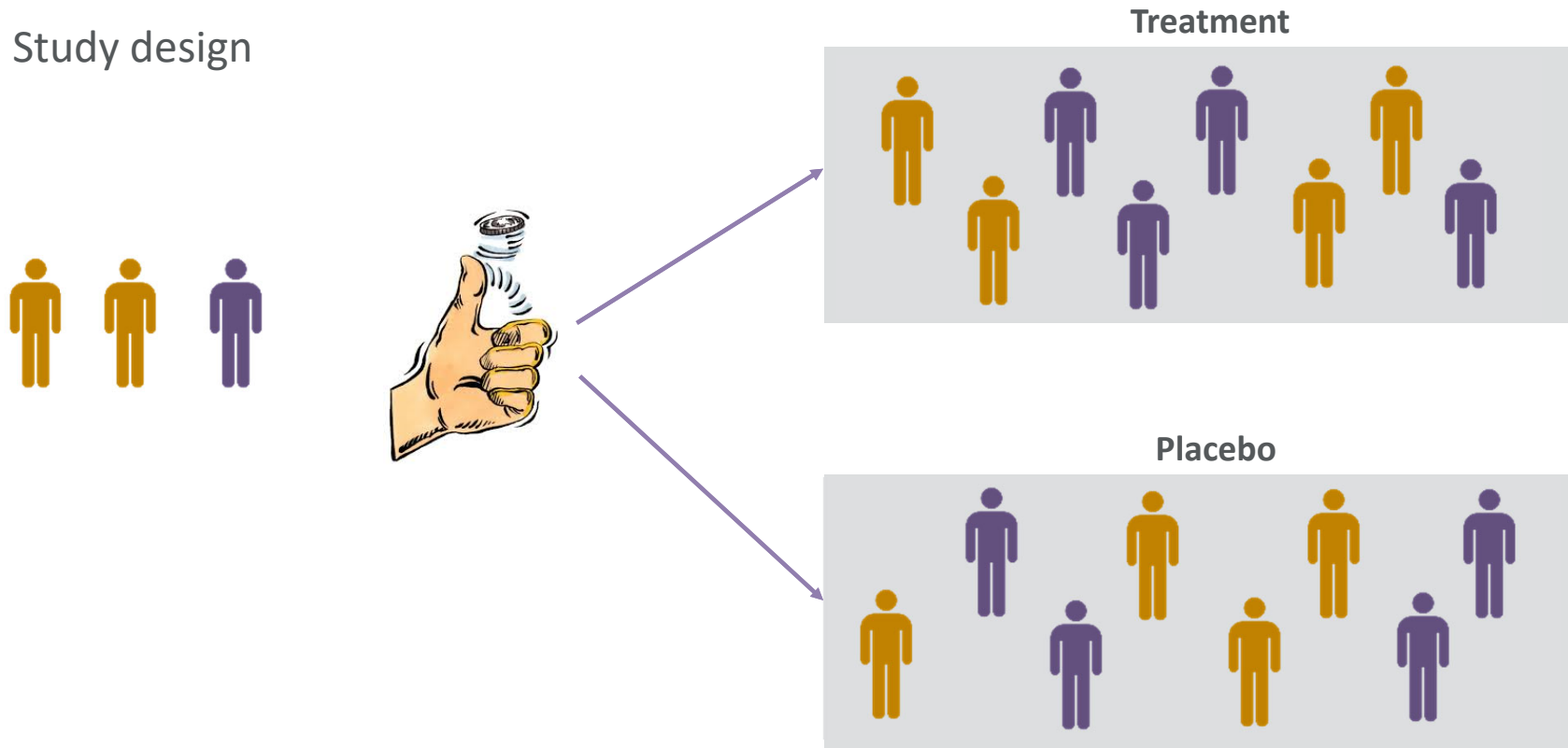
- Be transparent in reporting the research question/objective
- Be transparent in reporting methods
 - Describe all “candidate variables” considered rather than selectively reporting only those in final models (STROBE guidance)
 - Report all model selection procedures / criteria
- Be transparent in reporting results
 - Include confidence intervals as a measure of precision
 - Interpretations align with goal (explanatory vs. prediction)
- Acknowledge any limitations with methods or interpretations

Considerations in Randomized Controlled Trials



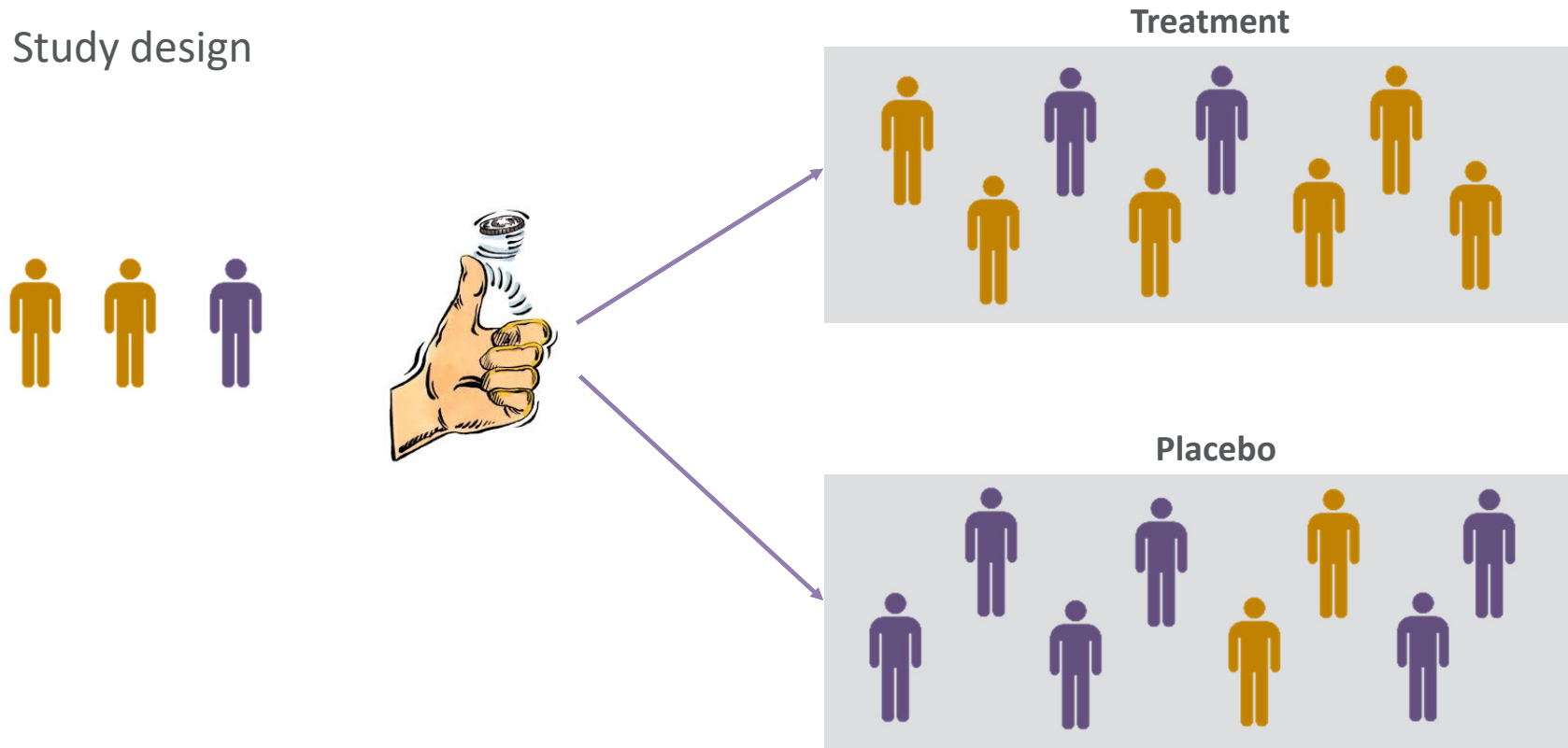
Why should I consider covariates?

- Study design



Why should I consider covariates?

- Study design



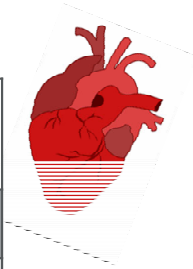
Why should I consider covariates?

- Study design
 - Incorporate covariates in the randomization scheme

Stratification



	Heart Disease	No Heart Disease
Male		
Female		



	Smoker		Non-Smoker	
	Heart Disease	No Heart Disease	Heart Disease	No Heart Disease
Male				
Female				

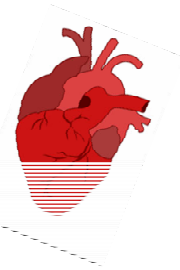
Why should I consider covariates?

- Study design
 - Incorporate covariates in the randomization scheme

Stratification



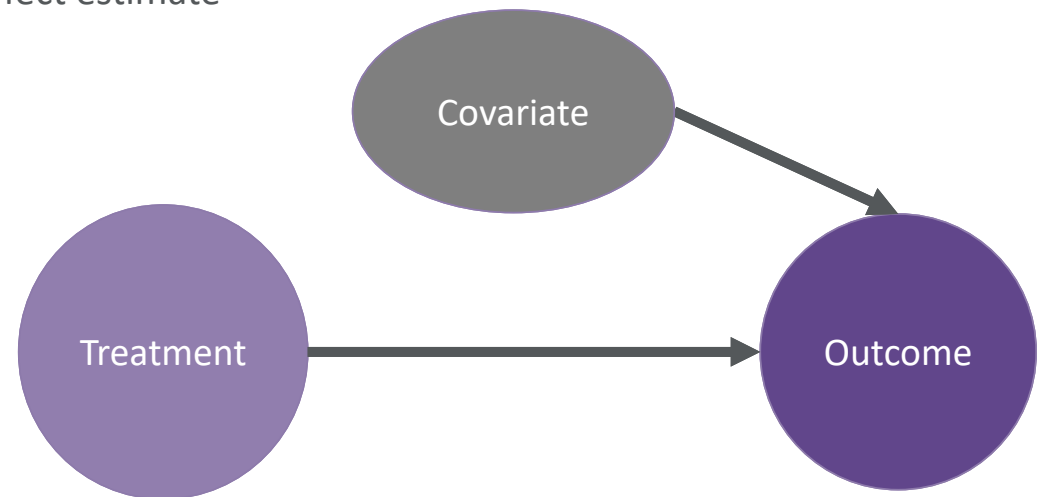
	Heart Disease	No Heart Disease
Male		
Female		



Covariate-Adaptive / Minimization

Why should I consider covariates?

- Analysis
 - Included in the design
 - Protection against chance imbalance
 - Increase power
 - Continuous outcomes: Improvement in precision of treatment effect
 - Binary / Time-to-event outcomes: Larger effect estimate



What covariates to consider?

- Believed / known to be prognostic of the outcome
- Measured at baseline (prior to randomization)
- Baseline measure of the outcome
- Center (site) in a multicenter study
- Exclude redundant variables

What not to do with baseline covariates?

- Formally test for ‘balance’!
 - CONSORT: Differences in baseline characteristics are the result of chance rather than bias
 - P-values assess the probability that observed baseline differences could have occurred due to chance – but we already know that - superfluous

“The decision to adjust should not be determined by whether baseline differences are statistically significant”

- CONSORT 2010 explanation and elaboration

General Guidance for Study Planning



Study Design

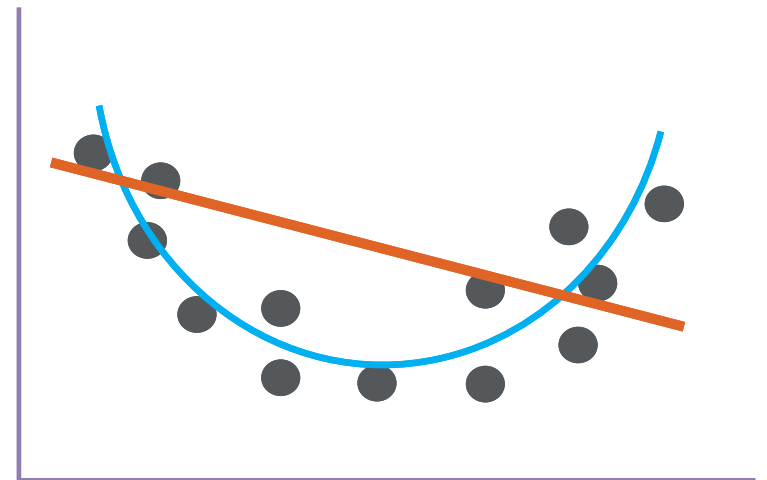
- Pre-specify planned adjustments
 - Methods for covariate adjustment / selection
- Consider the impact of covariate adjustment on power
 - Incorporate planned adjustments in sample size calculations

Prospective data collection

- Timing of measurements
- Measurement error
 - Validated measures
 - Consistency in collecting covariates
- Sources of variation in measurements
- Implement processes to support high quality, complete data collection

Additional covariate considerations

- Assessing functional form
 - Non-linear relationships
 - Diagnostic plots
- Categorizing / Dichotomizing
- Time-independent vs. time-dependent variables



“A good model should be substantively motivated, appropriate to the inferential goal and sample size, interpretable, and persuasive”

- Vittinghoff et al., *Regression Methods in Biostatistics*

Resources

- Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Medicine* 4(10): e297. <https://doi.org/10.1371/journal.pmed.0040297>
- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomized trials. *BMJ* 2010;340:c869 doi: 10.1136/bmj.c869
- FDA Guidance on Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products: Guidance for Industry. Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. (May 2021). Available at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products>
- Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. *Biom J.* 2018;60(3):431-449. doi:10.1002/bimj.201700067
- Vittinghoff E., Glidden D.V., Shiboski S.C., McCulloch C.E. (2012) Predictor Selection. In: Regression Methods in Biostatistics. Statistics for Biology and Health. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-1353-0_10
- Field-Fote, Edelle [Edee] PT, PhD, FAPTA Mediators and Moderators, Confounders and Covariates: Exploring the Variables That Illuminate or Obscure the “Active Ingredients” in Neurorehabilitation, *Journal of Neurologic Physical Therapy*: April 2019 - Volume 43 - Issue 2 - p 83-84 doi: 10.1097/NPT.0000000000000275
- Kahan, B.C., Jairath, V., Doré, C.J. et al. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials* 15, 139 (2014). <https://doi.org/10.1186/1745-6215-15-139>
- Ciolino, J.D., Palac, H.L., Yang, A. et al. Ideal vs. real: a systematic review on handling covariates in randomized controlled trials. *BMC Med Res Methodol* 19, 136 (2019). <https://doi.org/10.1186/s12874-019-0787-8>

Upcoming *Statistically Speaking* Lectures

Wednesday, January 12
12-1pm

Meta-analysis - When You Can Do It, When You Can't and What Info You Need

Lutfiyya N Muhammad, PhD, MPH, Assistant Professor, Division of Biostatistics, Department of Preventive Medicine

Thursday, February 3
12-1pm

Capturing Racial and Ethnic Data and Considerations for Analysis

Jody D Ciolino, PhD, Associate Professor, Division of Biostatistics, Department of Preventive Medicine

Thursday, March 3
1-2pm

Considerations when Leveraging Electronic Health Records for Causal Inference: A “Create-Your-Own-Data” Adventure

Lucia C Petito, PhD, Assistant Professor, Division of Biostatistics, Department of Preventive Medicine

All lectures will be held via Zoom: <https://northwestern.zoom.us/j/99300239887>

Thank you!

Biostatistics Collaboration Center

Search web or people

About Us ▾Research Services and SupportEducationPeople

Biostatistics Collaboration Center

REQUEST AN APPOINTMENT

Expertise in biostatistics, statistical programming and data management

Since 2004, the Biostatistics Collaboration Center (BCC) has partnered with Northwestern investigators at every level – from residents and postdoctoral fellows to junior faculty and well-established senior investigators.



Meet Our Team

Our faculty and staff provide biostatistical and health services information across Chicago and Evanston Northwestern clinical partners. Meet our team and learn more about our expertise.

OUR PEOPLE

Please note, we have been experiencing a high volume of requests and responses may be delayed. Your patience is appreciated as we work to field requests in as timely a manner as possible!



Northwestern
Feinberg School of Medicine

44