

All materials available at:

https://github.com/eleanormurray/NeuRA_Sydney_2019/

Getting the most out of pragmatic trials – beyond the intention-to-treat effect

Eleanor Murray, ScD

Department of Epidemiology



School of
Public Health

NeuRA Sydney

Nov 20-21, 2019

@EpiEllie

Outline: Today

1. What is a pragmatic trial?
2. Asking good causal questions
3. Understanding causal graphs
4. Intention-to-treat effects, but better
5. Per-protocol effects, done right
6. Guidelines for causal inference from pragmatic trials

Outline: Tomorrow

Practicum: Estimating causal effects from pragmatic trials with survival outcomes

- Make sure you have R, SAS, or Stata installed on your computer.
 - If using R, install required packages tonight in case of WIFI issues
- Link to materials available – check that you can access them before you leave today!
 - https://github.com/eleanormurray/NeuRA_Sydney_2019/

Part I: What is a pragmatic trial?



RCTs are seen as optimal for detecting treatment effects

Randomization prevents confounding ...

But, results can have limited applicability for clinical decision-making

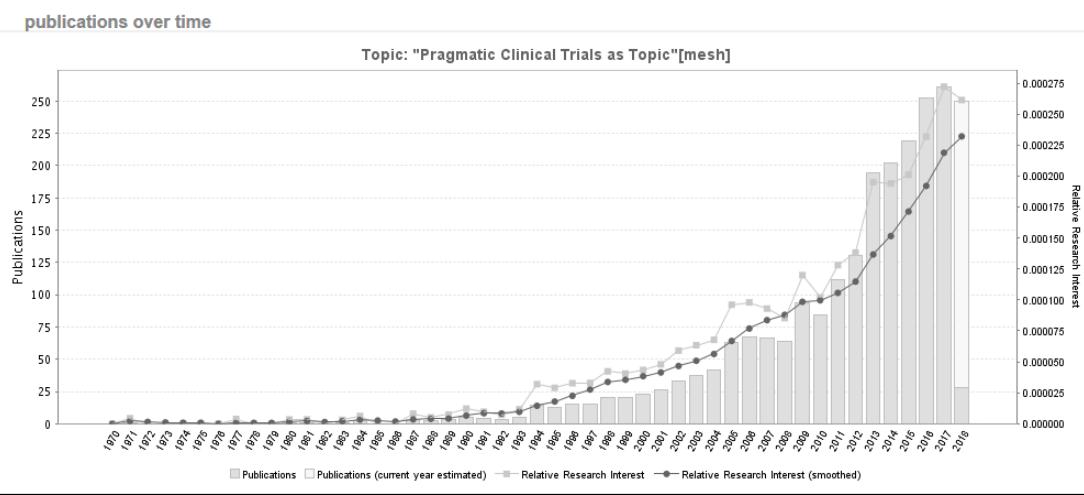
- highly selected population
- short duration & intermediate or surrogate outcomes
- comparators not clinically relevant

Solution: Pragmatic randomized trials

Definition: A randomized trial designed to assess real-world effectiveness of interventions

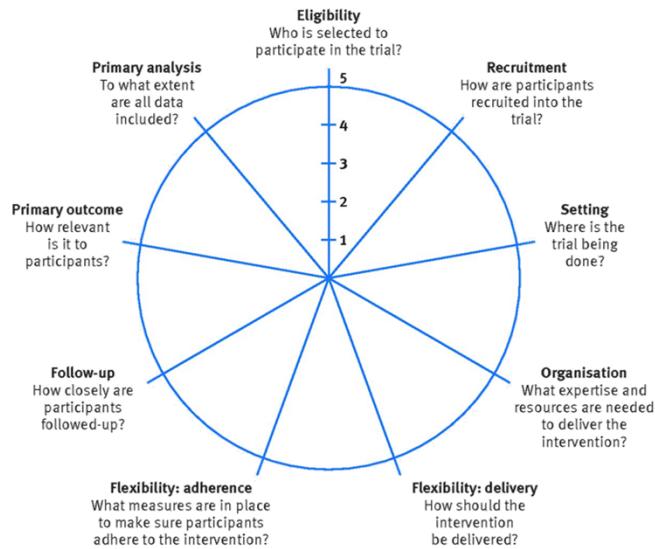
Pragmatic randomized trials are designed to ask clinically relevant, generalizable, questions

Pragmatic randomized trials



[https://www.gopubmed.org/web/gopubmed/statistics/"Pragmatic+Clinical+Trials+as+Topic"\[mesh\]](https://www.gopubmed.org/web/gopubmed/statistics/)

PRECIS-II: Pragmatic-Explanatory Continuum Indicator Summary 2



Loudon et al. BMJ 2015;350:h2147

CONSORT extension

Section	Item	Standard CONSORT description	Extension for pragmatic trials
Title and abstract	1	How participants were allocated to interventions (eg, "random allocation," "randomised," or "randomly assigned")	
Background	2	Scientific background and explanation of rationale	Describe the health or health service problem that the intervention is intended to address and other interventions that may commonly be aimed at this problem
Participants	3	Eligibility criteria for participants; settings and locations where the data were collected	Eligibility criteria should be explicitly framed to show the degree to which they include typical participants and/or, where applicable, typical providers (eg, nurses), institutions (eg, hospitals), communities (or localities eg, towns) and settings of care (eg, different healthcare financing systems)
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered	Describe extra resources added to (or resources removed from) usual settings in order to implement intervention. Indicate if efforts were made to standardise the intervention or if the intervention and its delivery were allowed to vary between participants, practitioners, or study sites. Describe the comparator in similar detail to the intervention
Objectives	5	Specific objectives and hypotheses	

Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D for the CONSORT and Pragmatic Trials in Healthcare (PractiHC) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008; 337:a2390.

CONSORT extension

Section	Item	Standard CONSORT description	Extension for pragmatic trials
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (eg, multiple observations, training of assessors)	Explain why the chosen outcomes and, when relevant, the length of follow-up are considered important to those who will use the results of the trial
Sample size	7	How sample size was determined; explanation of any interim analyses and stopping rules when applicable	If calculated using the smallest difference considered important by the target decision maker audience (the minimally important difference) then report where this difference was obtained
Randomisation—sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (eg, blocking, stratification)	
Randomisation—allocation concealment	9	Method used to implement the random allocation sequence (eg, numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned	
Randomisation—implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups	

CONSORT extension

Section	Item	Standard CONSORT description	Extension for pragmatic trials
Blinding (masking)	11	Whether participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment	If blinding was not done, or was not possible, explain why
Statistical methods	12	Statistical methods used to compare groups for primary outcomes; methods for additional analyses, such as subgroup analyses and adjusted analyses	See Part 6: Guidelines
Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended)—specifically, for each group, report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome; describe deviations from planned study protocol, together with reasons	The number of participants or units approached to take part in the trial, the number which were eligible, and reasons for non-participation should be reported
Recruitment	14	Dates defining the periods of recruitment and follow-up	
Baseline data	15	Baseline demographic and clinical characteristics of each group	
Numbers analysed	16	Number of participants (denominator) in each group included in each analysis and whether analysis was by “intention-to-treat”; state the results in absolute numbers when feasible (eg, 10/20, not 50%)	

CONSORT extension

Section	Item	Standard CONSORT description	Extension for pragmatic trials
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (eg, 95% CI)	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating which are prespecified and which are exploratory	
Adverse events	19	All important adverse events or side effects in each intervention group	
Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes	
Generalisability	21	Generalisability (external validity) of the trial findings	Describe key aspects of the setting which determined the trial results. Discuss possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial
Overall evidence	22	General interpretation of the results in the context of current evidence	

What do pragmatic trials look like: A Literature review

Search criteria: All pragmatic randomized trials published from Sept 2006 to Sept 2016 in:

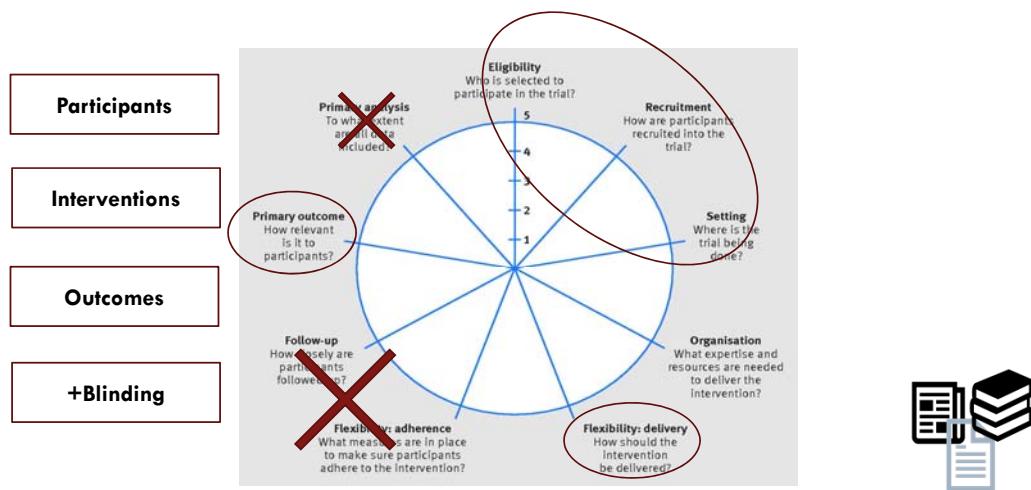
- New England Journal of Medicine
- Journal of the American Medical Association
- British Medical Journal
- The Lancet

Search terms for:

- A. Randomized trials
- B. Characteristics of pragmatic trials
- C. Phase IV trials

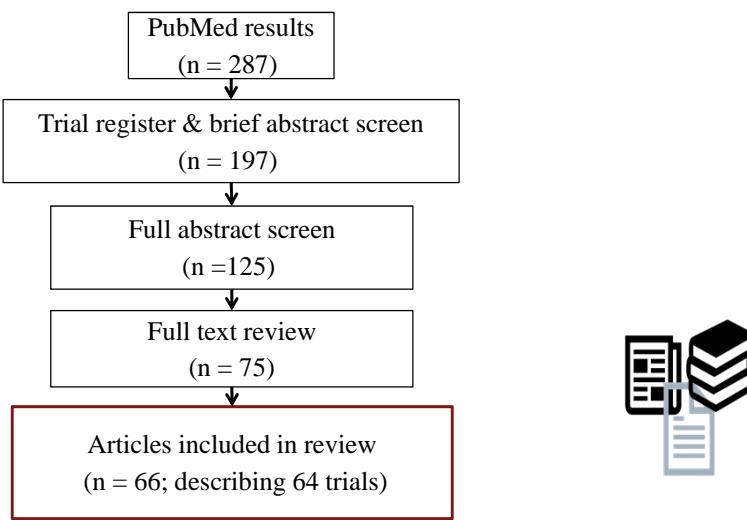


How to identify a pragmatic trial: Modified PRECIS-II



Kirsty Loudon et al. BMJ 2015;350: bmj.h2147

Results: 63 pragmatic trials over 10 years



Characteristics of recent pragmatic trials

Characteristic	Trials (N = 64)
Number of Participants: median (IQR)	511 (241, 822)
Length of follow-up: median (IQR)	12 months (6, 24)
% women: median (IQR)	55% (40, 64)
Number of treatment arms:	
2	39 (62%)
3	14 (22%)
4 or more	10 (16%)
Trial objective:	
Superiority	53 (82%)
Non-inferiority	8 (13%)
Both	2 (3%)

Characteristics of pragmatic trials

Characteristic	Trials (N = 64)
Intervention type:	
Medication	18 (29%)
Surgery protocol	13 (21%)
Medical device	12 (19%)
Treatment protocol	8 (13%)
Counselling or therapy	5 (8%)
Diagnostic test	4 (6%)
Other	4 (6%)
Specialty:	
Cardiovascular disease	8 (13%)
Infectious disease	7 (11%)
Reproductive health	7 (11%)
Primary care	7 (11%)
Orthopedics	6 (10%)
Psychiatry	5 (8%)
Other	23 (37%)

Trade-off: clinical relevance vs bias

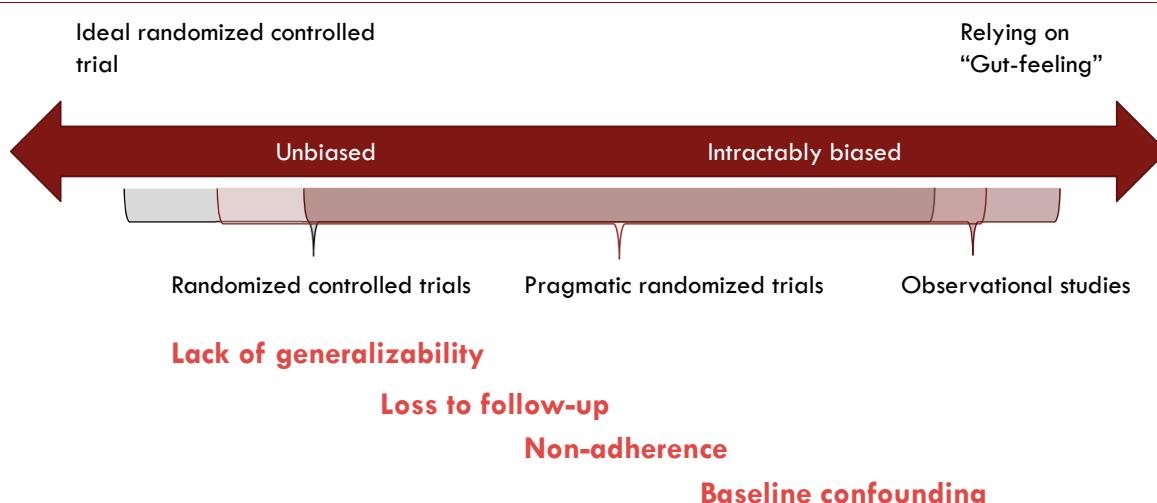
Long duration

- + Allows a more relevant primary outcome
- Increases chances of loss to follow-up
- Increases chances of non-adherence

Usual care as comparison

- + Answers a more clinically relevant question
- Lots of variation in care received
- Potential bias in ITT **away** from null

A bias continuum



How well do pragmatic trials address bias?

Characteristic	Trials (N = 64)
Effect estimated:	
Intention-to-treat only	41 (65%)
+ Per-protocol	22 (35%)
Method for per-protocol analysis (n=22)	
Per-protocol population	20 (91%)
Statistical adjustment for non-compliance	1 (4%)
Other ("Landmark")	1 (4%)
Method for loss to follow-up (n=31)	
Complete case only	31 (86%)
Multiple imputation	7 (13%)
Complete case with sensitivity analysis	6 (11%)
Worst-case imputation	3 (5%)
Last observation carried forward / linear interpolation	3 (5%)
Best case imputation	1 (2%)
Other, unclear, or not reported	4 (7%)

Not Well!

What is the key challenge for validity in pragmatic trials?

Post-randomization events that are affected by prior treatment and affect future treatment

- Non-adherence
- Loss to follow-up

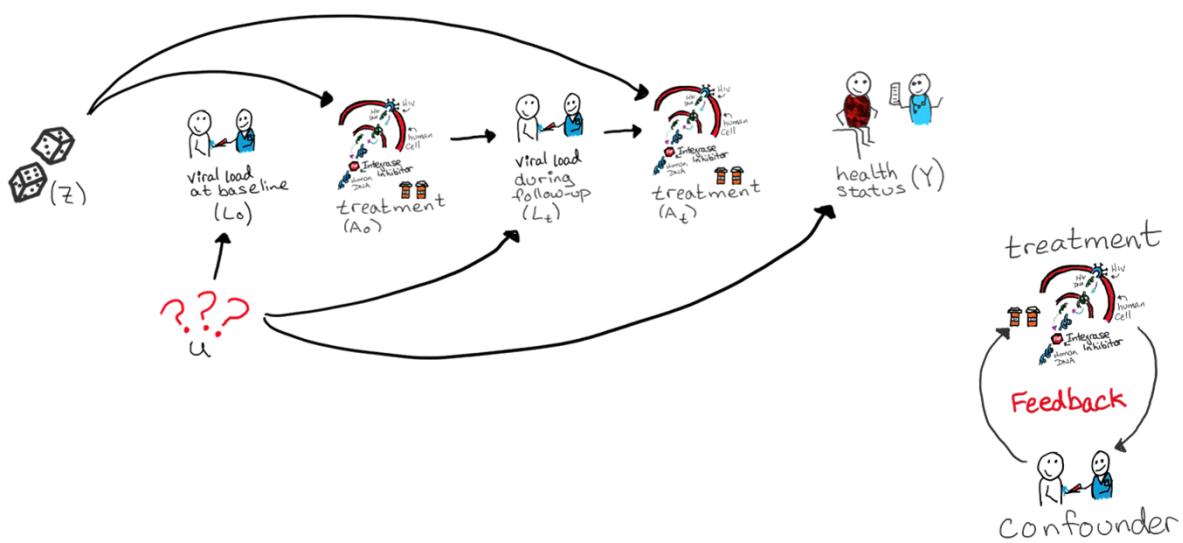
Solution: Only the **g-methods** can handle feedback between time-varying treatments & time-varying confounders.

A quick handshake intro to g-methods (more later!)

1. Inverse probability weighting (aka IPW) of marginal structural models
2. (Parametric) G-formula
3. Doubly-robust estimation (aka targeted maximum likelihood estimation or TMLE if estimated using machine learning)
4. G-estimation of structural nested models



G-methods generalize estimation to treatment-confounder feedback



G-methods are roughly similar to ...

Use when
you have
treatment-
confounder
feedback,
and

Inverse probability weighting	\approx	Propensity scores
(Parametric) G-formula	\approx	Standardization
G-estimation	\approx	Instrumental variables

... you
would
normally
use these.

Summary: Part I what are pragmatic trials?

- Randomized trials designed to answer clinical practice questions
- Designed to include more representative patients, settings, and health conditions
- Typically unblinded
- Vulnerable to observational data-type problems: post-randomization confounding by non-adherence and loss to follow-up

Discussion time!

Has anyone here ever been a part of a pragmatic trial?

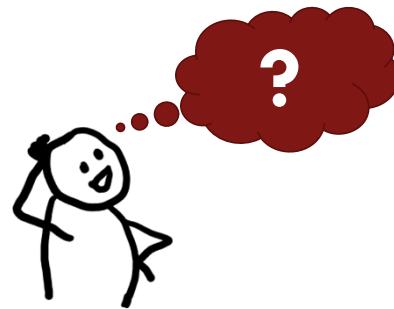
- Why did you choose this design? What challenges did you face? How did you deal with loss to follow-up & non-adherence?



Where to get more information: Part I pragmatic trials

- Murray et al. 2018. J Clin Epi 103:10-21.
- Loudon et al. BMJ 2015;350:h2147
- Zwarenstein M, et al for the CONSORT and Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008; 337;a2390.

Part II: Asking good causal questions



What is causal inference?

Not about identifying causes!

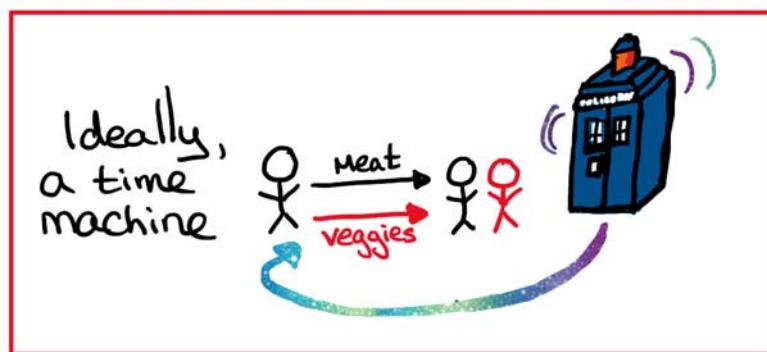


What is causal inference?

Estimating the strength of cause and effect relationships!

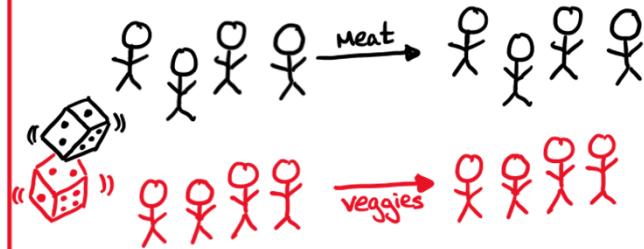


What do we want to know?

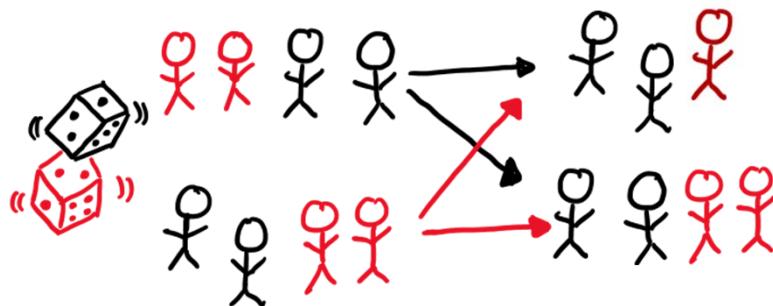


What do we want to know?

Next best thing,
a randomized
controlled trial

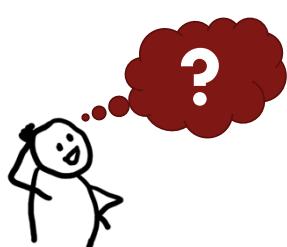


But real life is messy...

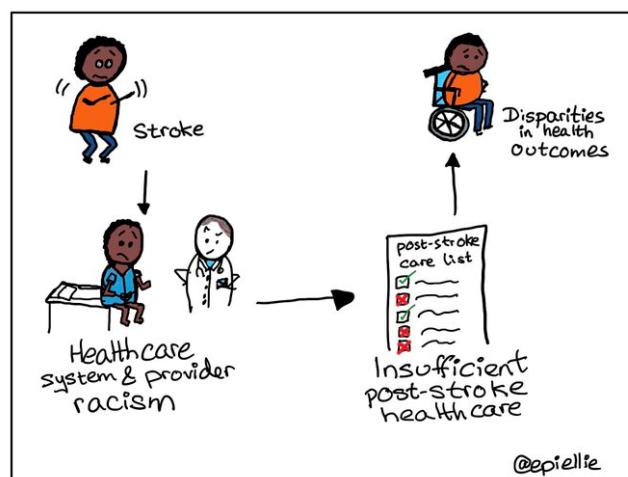


How do we ask good causal questions?

Specify a clear protocol for exposure including how and when it happens!



Defining exposure can be hard...



What makes an exposure complex?

Multiple components

- e.g. race, socio-economic position, cognitive behavioral therapy

Interference between individuals

- e.g. infections, behaviors & habits, education

Exposures that vary over space

- e.g. air pollution, access to goods & services

Exposures that vary over time

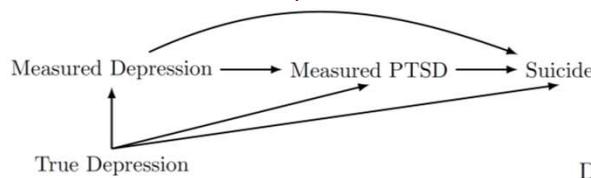
- e.g. medication usage, unhealthy habits

Simple exposures that could occur at any time

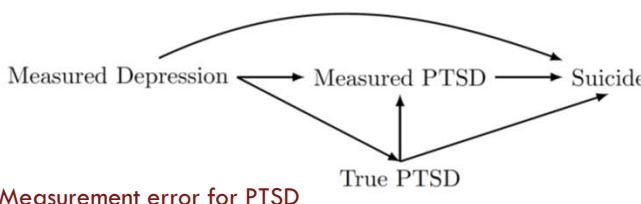
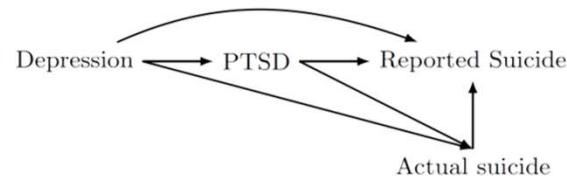
- e.g. surgery

...and measuring exposure can be even harder

Measurement error for depression

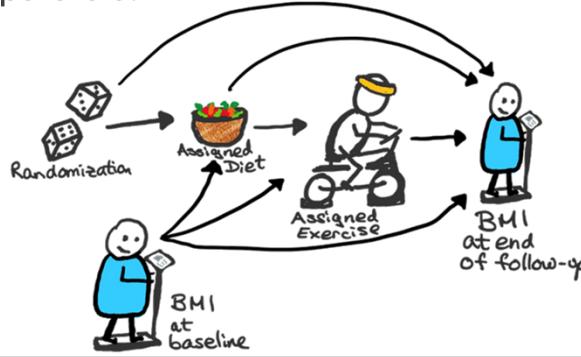


Measurement error for suicide



Complex exposures give us complicated answers, even when we can intervene...

- Do we always need to give all parts of the intervention?
- Does the timing of the intervention or pieces matter?
- How would the intervention work in groups with other types of usual care/ comparators?



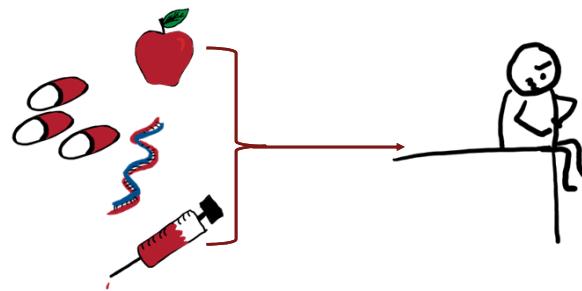
... so what about when we can't intervene?

We need to be even more careful about the questions we ask!

Remember: in a pragmatic trial we intervene to ASSIGN exposure, but we can't intervene to FORCE exposure

Why are well-defined causal questions important for complex exposures?

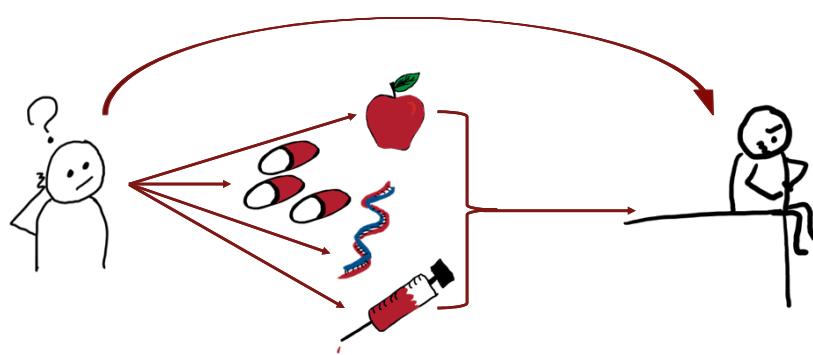
- When there are multiple possible ‘interventions’ and we don’t specify one, our answer is a weighted average of all ‘interventions’ but we don’t know the weights



Murray, 2016. Agent-based models for causal inference. Harvard University.
Murray, et al. 2019. Medical Decision Making [in press]

Why are well-defined causal questions important for complex exposures?

- Worse, if the ‘intervention’ is ill-defined, the confounding is probably also ill-defined!



Murray, 2016. Agent-based models for causal inference. Harvard University.
Murray, et al. 2019. Medical Decision Making [in press]

Asking questions that matter for complex exposures

- What is the intervention you would do *if* you could do an experiment?
 - How would that experiment help you make treatment, policy, or other decisions?
- Are we asking a *specific enough* question to get an answer we can understand and act upon?

The best questions lead to action



What would the difference in total income be for all tipped workers:

- if all tipped workers had been women versus
- if all **those same** tipped workers had been men?

➤ How do we **act** upon the answer to that question?

The best questions lead to action



What would the difference in total income be for women who work for tips:

- if all women received the amount of tips they typically receive versus
 - if those same women received the **amount of tips** that men typically receive?
- This question could help us **plan** policy interventions

Asking questions in pragmatic trials

- What is the treatment, policy, or other decision you want to make?
- Who do you want to make this decision for?
- When does the decision happen?

Patient-centered causal effects

Increasing interest in **shared decision-making** between patients and clinicians

- Lots of research on how to communicate information to patients
- But, don't know what types of causal information will help patients with these decisions

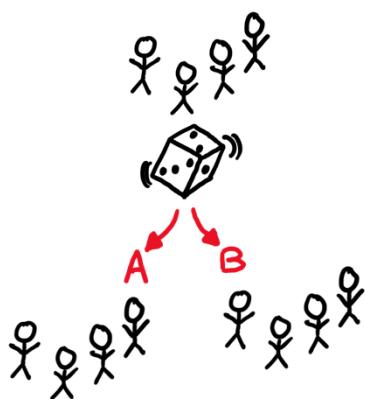
Patient-centered causal effects are causal effects which provide patients with usable information for shared decision-making

Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

What are our options for causal effects?

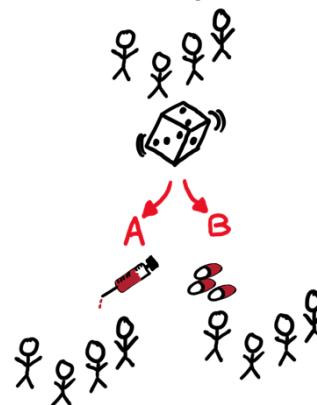
Intention-to-treat effects

- Effect of randomization to A vs B



Per-protocol effects

- Effect of receiving A vs B



Case Study: Intention-to-treat vs. per-protocol effect in the Women's Health Initiative

Women's Health Initiative (WHI) goal:

Estimate the health effects of postmenopausal estrogen plus progestin hormone therapy versus placebo.

Intention-to-treat effect and per-protocol effect

Per-protocol effect made the simplifying, but plausible, assumption that the number of women with contraindications for hormone therapy during the follow-up was negligible.

Rossouw JE et al. JAMA. 2002 Jul;288(3):321–33.
The Women's Health Initiative Study Group. Control Clin Trials. 1998 Feb;19(1):61–109.
Toh S et al. Epidemiology. 2010 Jul;21(4):528–39.

Case Study: Intention-to-treat vs. per-protocol effect in the Women's Health Initiative

- The estimated intention-to-treat hazard ratio of breast cancer was 1.25 (95% CI: 1.01, 1.54).
- The estimated per-protocol hazard ratio of breast cancer was 1.68 (95% CI: 1.24, 2.28).

Rossouw JE et al. JAMA. 2002 Jul;288(3):321–33.
The Women's Health Initiative Study Group. Control Clin Trials. 1998 Feb;19(1):61–109.
Toh S et al. Epidemiology. 2010 Jul;21(4):528–39.

Which would you rather be told as a patient?

- Assignment to hormone therapy increases the hazard of breast cancer by 25%
 - If taken exactly as happened in the trial, hormone therapy increases the hazard of breast cancer by 25%.
- If taken continuously, hormone therapy would increase the hazard of breast cancer by 68%!



Rossouw JE et al. JAMA. 2002 Jul;288(3):321–33.
The Women's Health Initiative Study Group. Control Clin Trials. 1998 Feb;19(1):61–109.
Toh S et al. Epidemiology. 2010 Jul;21(4):528–39.

Case Study: Characterizing patient-centered causal effects

Two medications, a standard and a new one, are equally effective

- Do you have a preference?

Some additional information is provided about adherence issues for one medication

- Do you have a preference?

A “per-protocol” argument is made in favor of the medication with low adherence

- Do you have a preference?



Murray EJ, et al. Journal of Clinical Epidemiology. 2018; 103: 10-21.

Intention-to-treat effects as starting point

But, patients really wanted detailed effects stratified by wide range of baseline characteristics

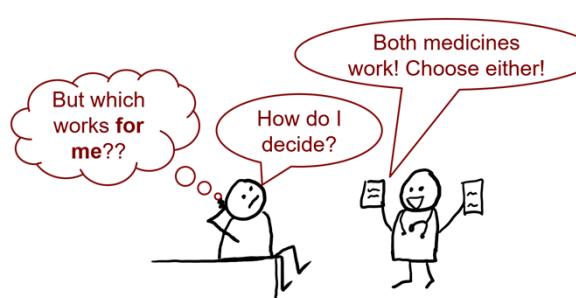
- “personalized-ish” medicine
- *How will this work for someone like me, taking the medications I am taking, and with the other conditions I have?*
- demographics, other diseases, disease severity, lifestyle characteristics, other medication use

“...also I would want to see the different groups, ethnic, race, sexes, weight, age, **the whole spiel**. I'd want to see all that first.”



Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

Intention-to-treat & stratified ITT are patient-centered



1. Patients want “personalized-ish” medicine
 - ❖ Stratified ITT effects are patient-centered
 - ✓ Involve patients & advocates in choosing *a priori* strata



Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

Before we could talk about adherence...

....Participants emphasized **strong** preference for clear superiority of one medication versus another

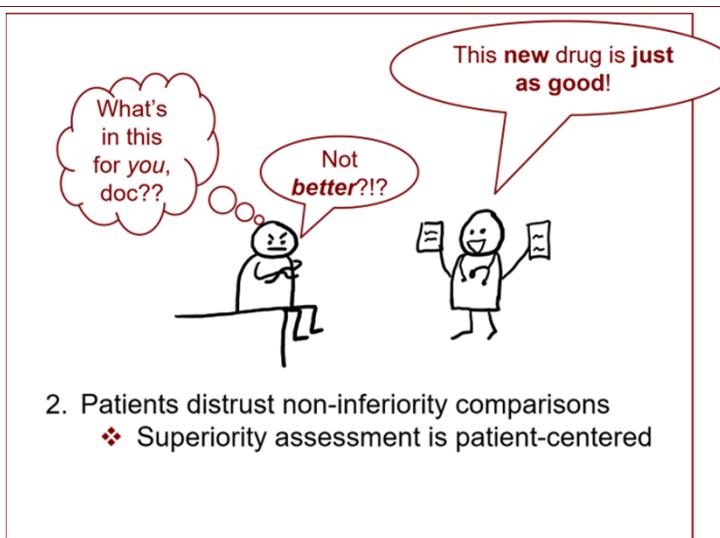
Non-inferiority information was viewed with extreme skepticism

"To me there has to be a point that they developed this drug. Like what else is going on with the drug?... **It's really not effective.** And not only is it **equally effective** than the drug that has already been around, it's inconvenient"



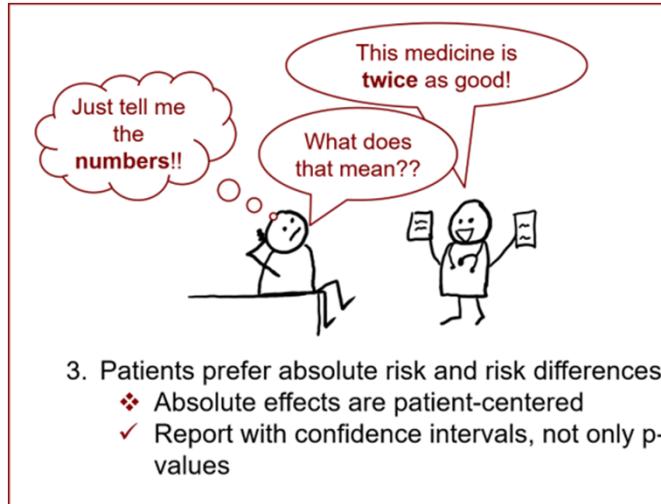
Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

Superiority estimates are patient-centered



Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

Patients understand absolute measures



3. Patients prefer absolute risk and risk differences

- ❖ Absolute effects are patient-centered
- ✓ Report with confidence intervals, not only p-values



Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

Per-protocol effects matter when patients **expect** to adhere

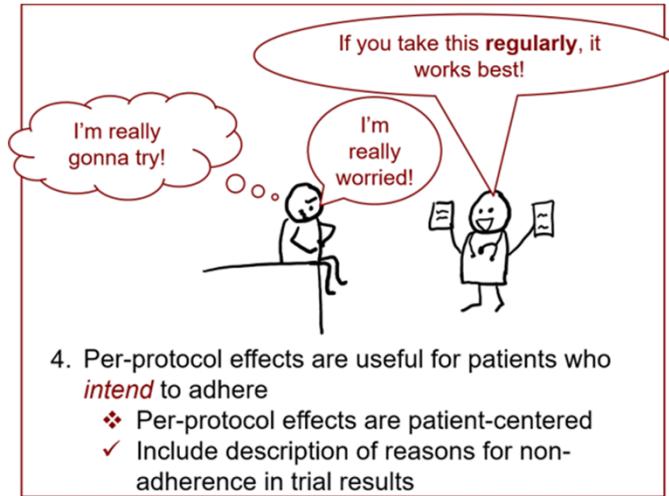
- “Some doctors believe that had all the patients taken the [standard/new] medication as prescribed, fewer patients would have had heart attacks compared to patients taking the [other, less convenient] medication”

“It would depend on how critical the case was. If I had serious COPD and there was, both parents had died of it, I would say, ‘You know what? I am **committed to my health**. I’m **committed to taking it as prescribed**.’ So I’d be **willing to try** the new [less convenient] drug.”



Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

Per-protocol effects are patient-centered when patients expect to adhere



Murray EJ, et al. *Journal of Clinical Epidemiology*. 2018; 103: 10-21.

But what is adherence anyway?

It's not just one thing!

- Take treatment if appropriate
- Take treatment regardless of appropriateness
- Take treatment if appropriate and *on schedule*
- Take every dose of treatment
- Take most doses of treatment

Adherence is a **complex** exposure!

Case study: Candesartan in Heart Failure Morbidity and Mortality

Candesartan was assessed in an RCT as secondary prevention for death and hospitalization among heart failure patients

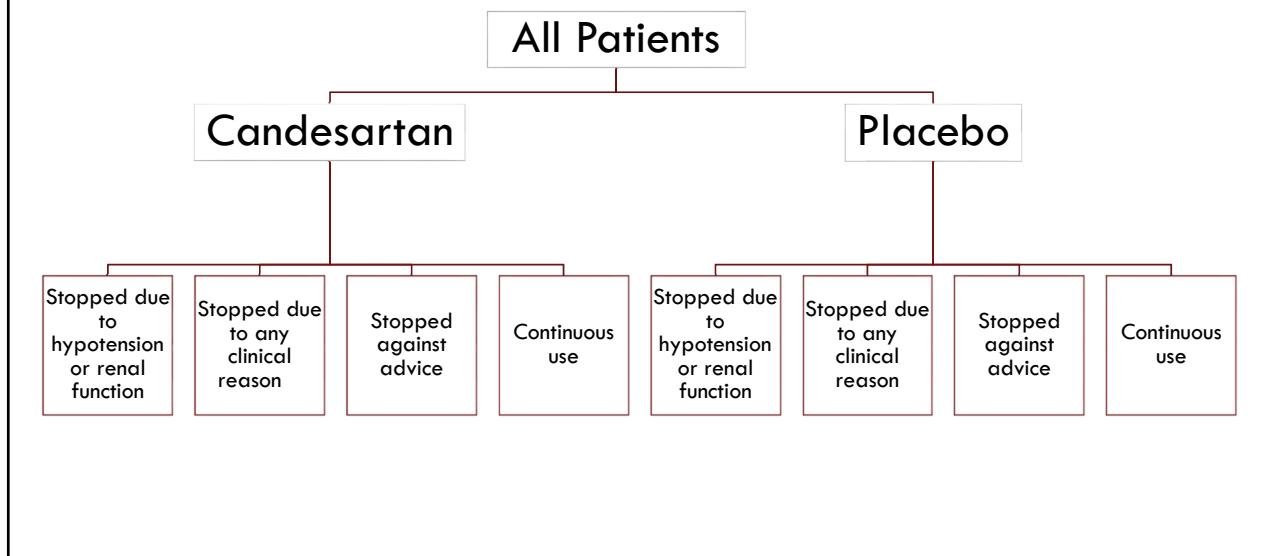
- Known complications:
 - Hypotension
 - Abnormal kidney function
- Potential for other, unknown, complications

Case study: Candesartan in Heart Failure Morbidity and Mortality

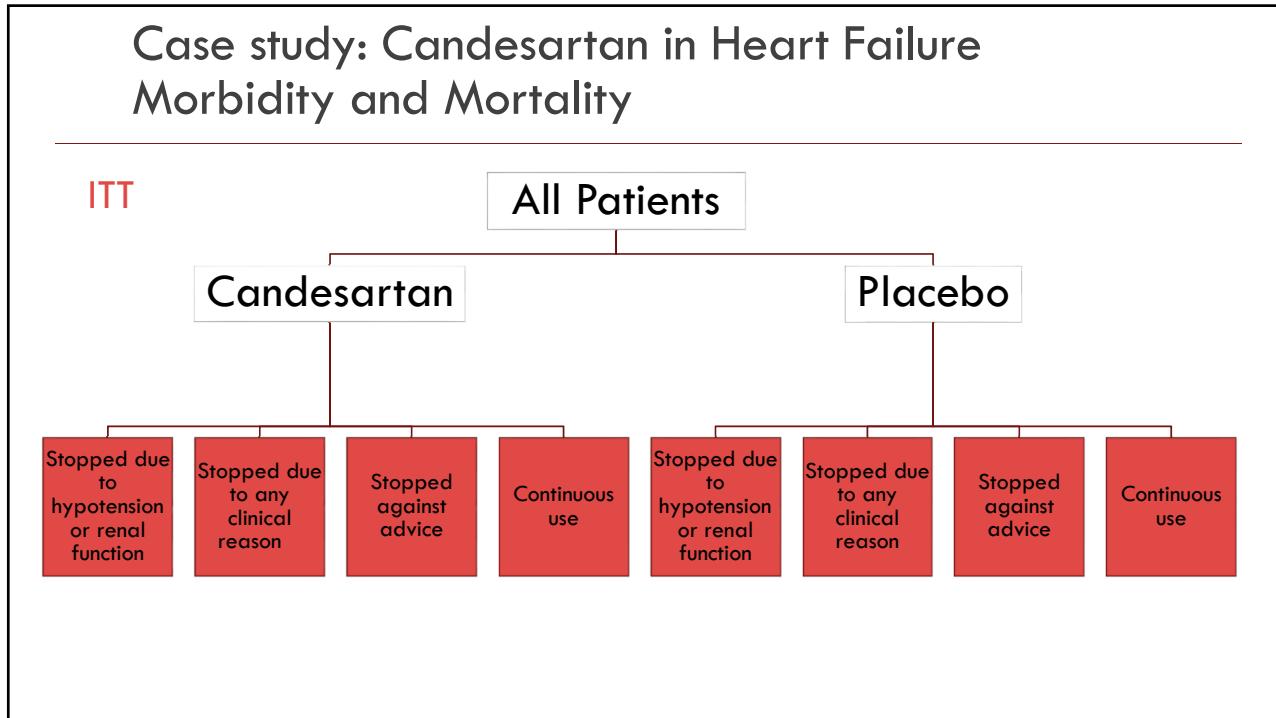
We estimated the per-protocol effect for TWO protocols:

- Take assigned medication (candesartan or placebo) at least 80% of the time, unless incident hypotension or abnormal renal function, after which follow clinician's instructions
- Take assigned medication (candesartan or placebo) at least 80% of the time, unless your treating clinician decides you should stop, after which follow clinician's instructions

Case study: Candesartan in Heart Failure Morbidity and Mortality



Case study: Candesartan in Heart Failure Morbidity and Mortality



Case study: Candesartan in Heart Failure Morbidity and Mortality

PPE 1

All Patients

Candesartan

Placebo

Stopped due
to
hypotension
or renal
functionStopped due
to any
clinical
reasonStopped
against
adviceContinuous
useStopped due
to
hypotension
or renal
functionStopped due
to any
clinical
reasonStopped
against
adviceContinuous
use

PPE 2

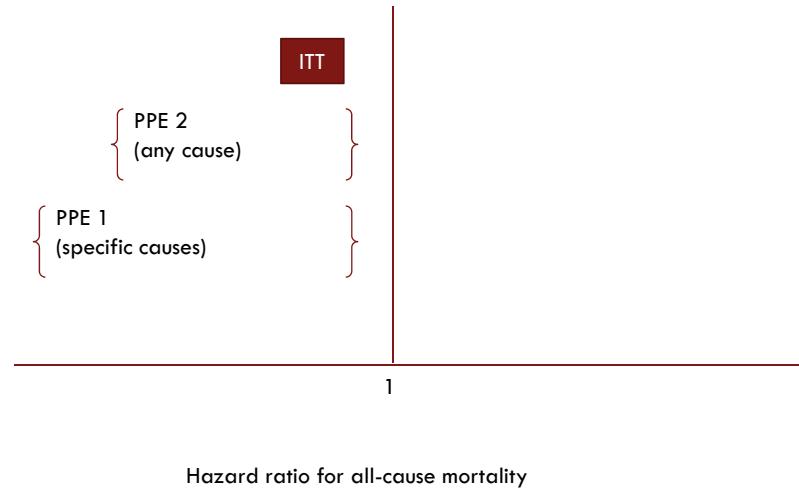
All Patients

Candesartan

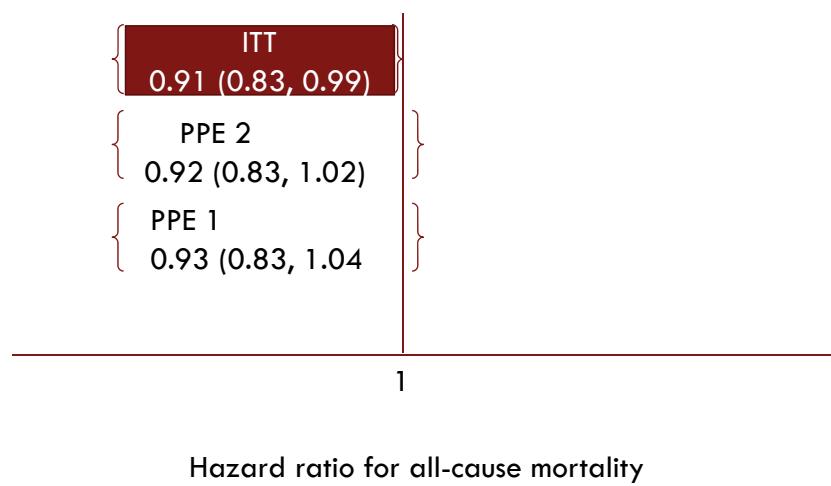
Placebo

Stopped due
to
hypotension
or renal
functionStopped due
to any
clinical
reasonStopped
against
adviceContinuous
useStopped due
to
hypotension
or renal
functionStopped due
to any
clinical
reasonStopped
against
adviceContinuous
use

What do we expect to see with these effects,
given that the trial is placebo controlled?



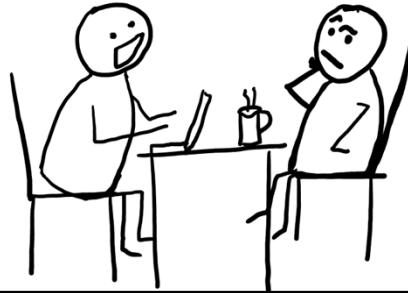
What do we actually see in the results?



Discussion time!

What tricky to define question have you faced in your research?

- When choosing questions, do you involve patients? Other stakeholders?
- Who is the target audience & what decision do they need to make?



Where to get more information: Part II good questions

- Murray et al. 2018. J Clin Epi 103:10-21.
- Murray EJ, et al. Adherence-adjustment in placebo-controlled randomized trials: an application to the Candesartan in Heart Failure randomized trial. 2019 [under review]
- Hernan MA. Annals of Epidemiology, 2016; 26(10):674-80.
- Rossouw JE et al. JAMA. 2002 Jul;288(3):321–33.
- The Women's Health Initiative Study Group. Control Clin Trials. 1998 Feb;19(1):61–109.
- Toh S et al. Epidemiology. 2010 Jul;21(4):528–39.

Morning Break



Part III: Understanding Causal Graphs



What are directed acyclic graphs (DAGs)?

DAGs are a graphical tool to represent relationships

Causal DAGs are a subset of DAGs which can help us design our studies, explain methods, and understand surprising results



When should we use causal DAGs?

Research that involves causal (or etiological) questions

Identify sources of systematic bias in study design and analysis

- Does confounding or selection bias exist?
- Is the causal effect of interest identifiable?

The structure of the bias can inform decisions about analytic approach

- What variables can be used to eliminate bias?
- Will traditional analytic methods suffice?

Understand unique or unexpected results

- Does this finding represent a true causal relationship or does it arise from bias?

Directed acyclic graphs have 3 basic rules

They must be:

- Graphs
- Directed
- Acyclic

Rule 1: What makes a graph?

Nodes



A

exposure

Edges



A — B

exposure → outcome

Pop-quiz: what elements are in this DAG?

Nodes:

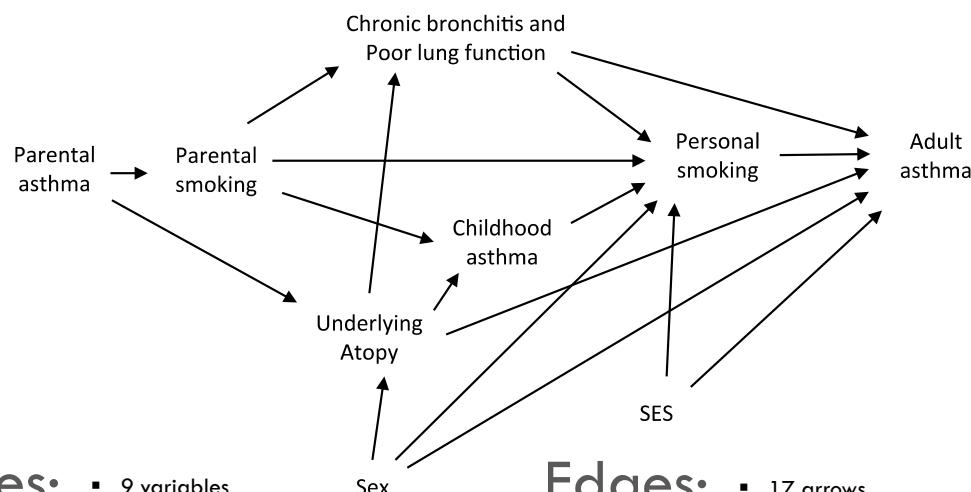
- Exercise
- Medication
- Blood pressure

Edges:

- Exercise → Medication
- Exercise → Blood pressure
- Medication → Blood pressure



DAGs for real research questions quickly get more complicated! What are the elements of this DAG?



Nodes: ▪ 9 variables

Edges: ▪ 17 arrows

Williamson EJ, et al. Introduction to causal diagrams for confounder selection. *Respirology*. 2014;19(3):303–311.

Rule 2: Edges must be *directed*



A may cause B



B may cause A



A doesn't cause B and B doesn't cause A

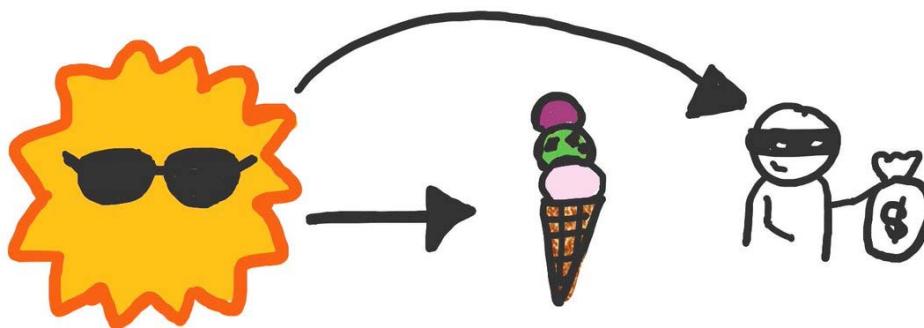
Causation happens from past to present or present to future, so edge directions should follow time

Edges:

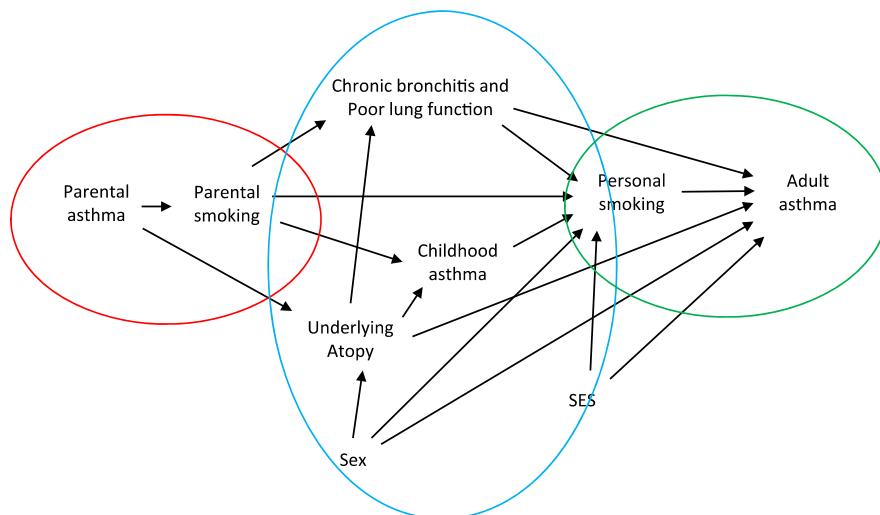
- Summer \rightarrow Ice cream
- Summer \rightarrow Crime

No edges:

- Ice cream \rightarrow Summer
- Crime \rightarrow Summer
- Ice cream \rightarrow Crime
- Crime \rightarrow Ice cream



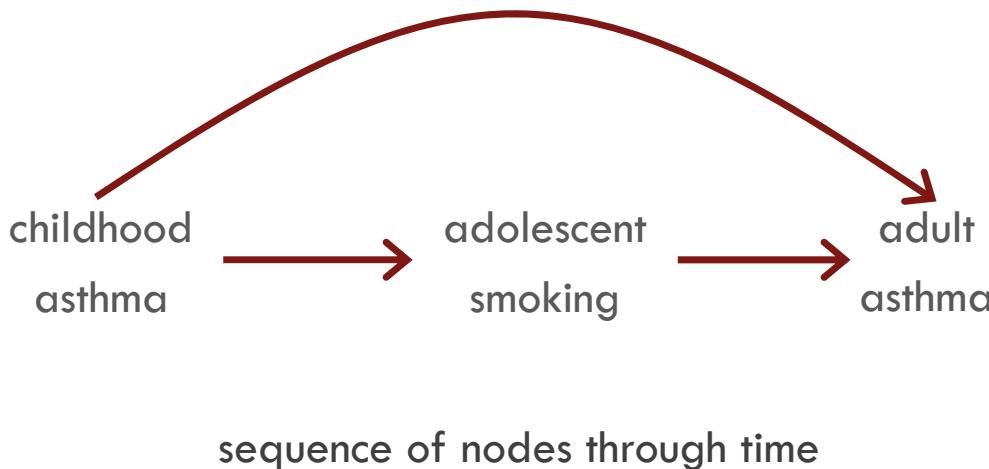
Parents before childhood before adulthood



Rule 3: DAGs must be acyclic. The future cannot cause the past!



So how do we capture dynamic relationships?



Causal directed acyclic graphs have 4 basic rules

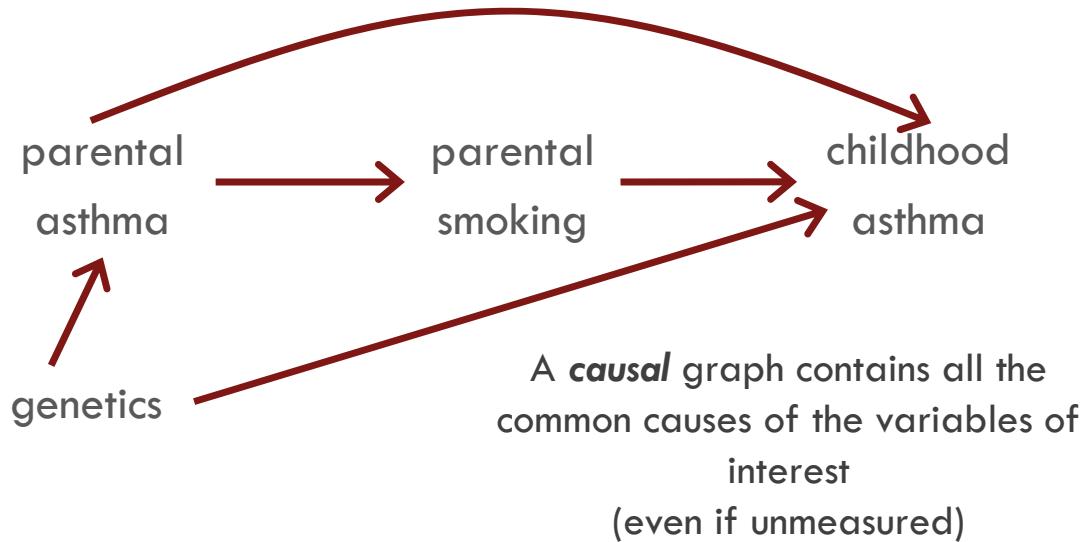
Rule 4: Causal DAGs must include

- All common causes of any pair of variables already on the graph

What variables should already be on the graph? Start with

- Exposure
- Outcome
- Restrictions

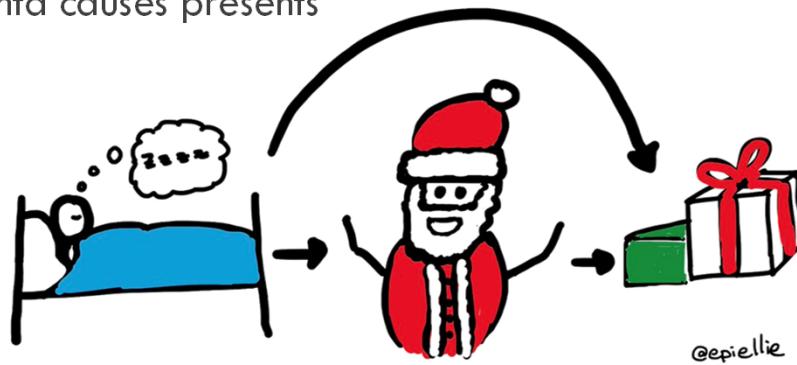
Smoking-asthma ***causal*** relationships



Directed edges encode causal assumptions

But the lack of an arrow is a **bigger assumption** than the presence of an arrow

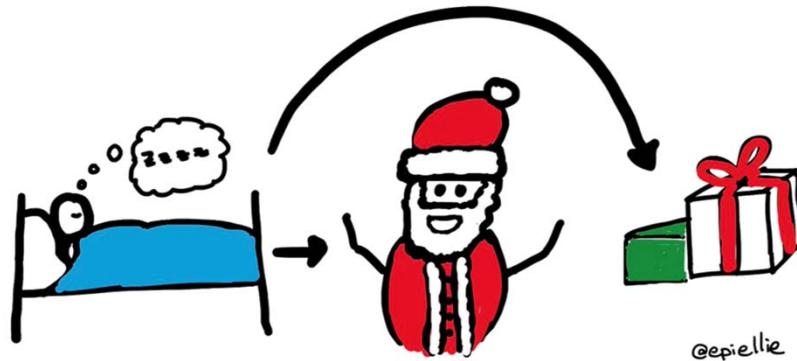
- The DAG below makes NO assumptions about whether Santa causes presents



Directed edges encode causal assumptions

But the lack of an arrow is a **bigger assumption** than the presence of an arrow

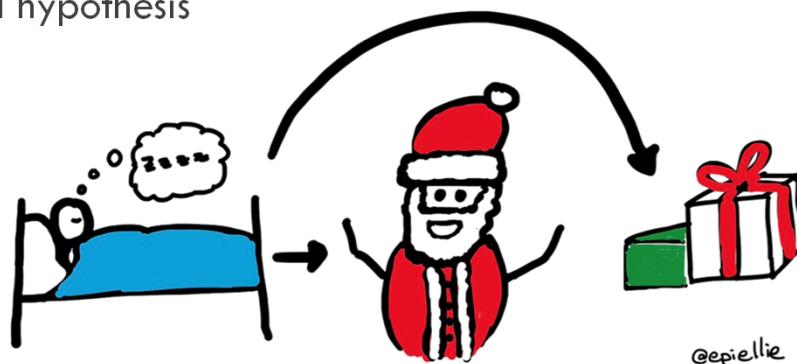
- The DAG below assumes Santa does NOT cause presents



Exception to this rule: DAGs under the *null*

If we want to estimate the causal effect of Santa on presents, we often draw the DAG under the null

- i.e. lack of an arrow from Santa to presents represents the null hypothesis



Nodes encode known and unknown variables



Nodes can also encode statistical adjustment or data restriction



How do we read graphs?

Sequences of connected nodes form *paths*



There are three basic types of paths

Chains:



Forks:



Colliders:



These paths are easier to read in *temporal order*

Chains:



Forks:



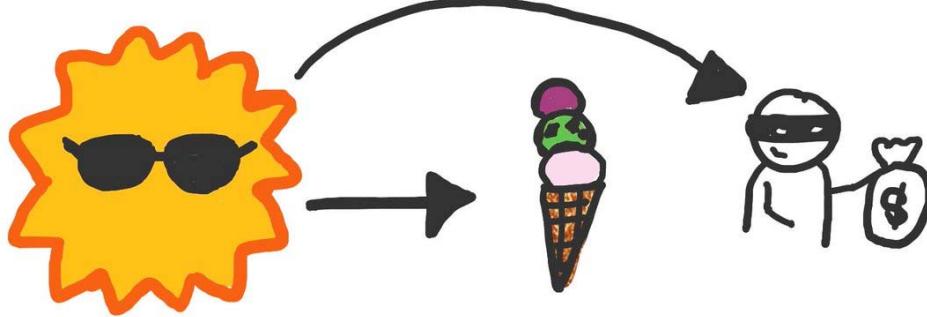
Colliders:



Chains imply causal relationships



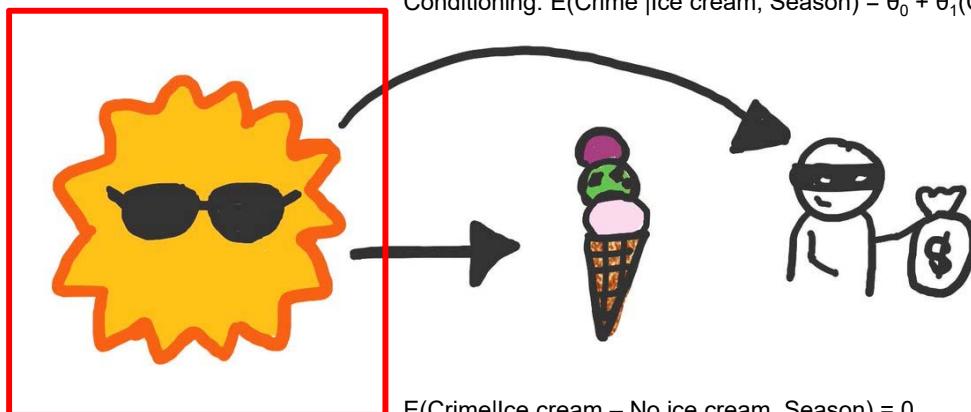
Forks can create spurious associations like confounding...



$$\begin{aligned} E(\text{Crime}|\text{Ice cream} - \text{No ice cream}) &\neq 0 \\ E(\text{Crime}|\text{Ice cream}) &= \gamma_0 + \gamma_1(\text{Crime}) \end{aligned}$$

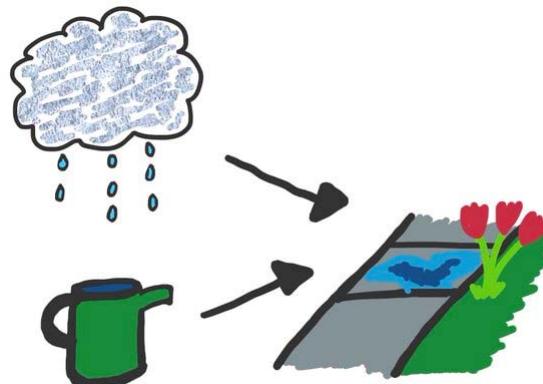
... we prevent this by restricting or conditioning

$$\begin{aligned} \text{Restriction: } E(\text{Crime}|\text{Ice cream, Season} = \text{summer}) &= \beta_0 + \beta_1(\text{Crime}) \\ \text{Conditioning: } E(\text{Crime}|\text{Ice cream, Season}) &= \theta_0 + \theta_1(\text{Crime}) + \theta_2(\text{Season}) \end{aligned}$$



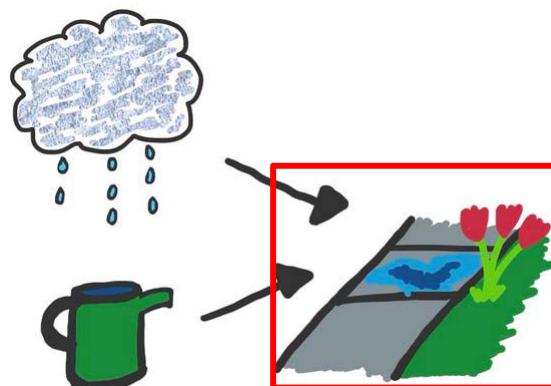
$$E(\text{Crime}|\text{Ice cream} - \text{No ice cream, Season}) = 0$$

Colliders imply no association



$$\begin{aligned} E(\text{Watering flowers} | \text{Rain} - \text{No rain}) &= 0 \\ E(\text{Watering flowers} | \text{Rain}) &= \gamma_0 + \gamma_1(\text{Rain}) \end{aligned}$$

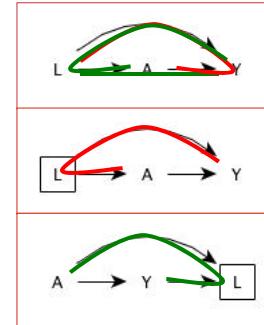
...but we can create associations by conditioning or restricting!



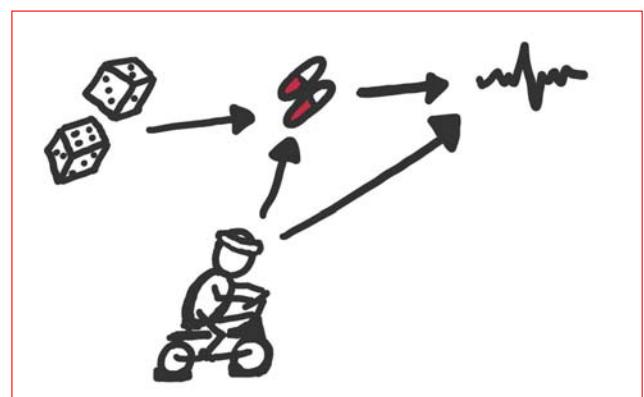
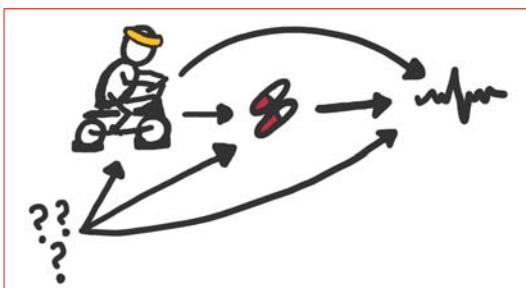
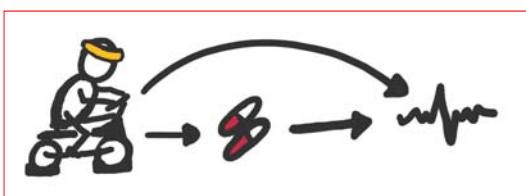
$$E(\text{Watering flowers} | \text{Rain} - \text{No rain, wet sidewalk}) \neq 0$$

Recap of the basics of DAGs

1. Missing arrows mean we assume no causation
2. Causation flows in the direction of the arrows only
3. If two arrow tails, or a tail & a head meet, association can flow between them *in both directions*
4. If two arrow heads meet, association is blocked
5. If we restrict or control on a variable, rules 3 & 4 are reversed



Some examples of DAGs for common biases



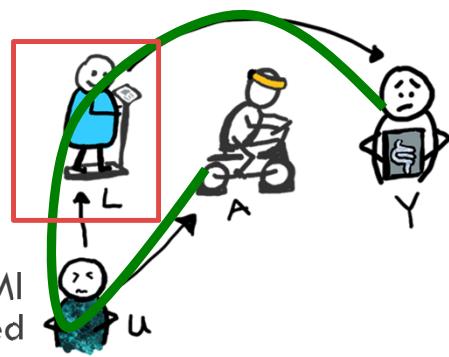
Reverse causation

- Y=Colon cancer diagnosis
- A=Physical activity
- L=BMI
- U=unmeasured preclinical disease

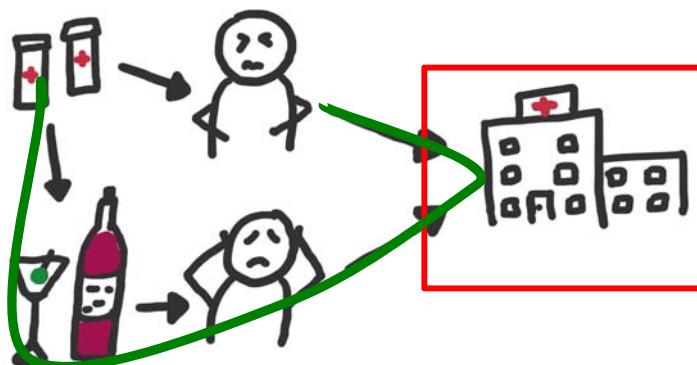
A and Y are confounded by pre-clinical disease

A and Y are independent conditional on BMI

- Backdoor criterion is met (all paths are blocked conditional on L)
- Adjusting for BMI allows for an unbiased estimate of the causal effect of physical activity on risk of colon cancer



Selection bias, aka Berkson's bias



- If we select controls from the hospital, we may induce association between the exposure and outcome because controls are hospitalized for a reason

Confounding with Time-varying Treatment

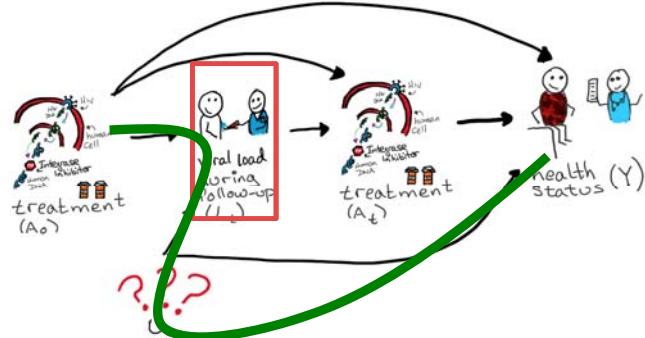
Notation: $A(0), A(1), \dots, A(t)$
denotes treatment at time points $0, 1, \dots, t$

Should we condition on $L(1)$?

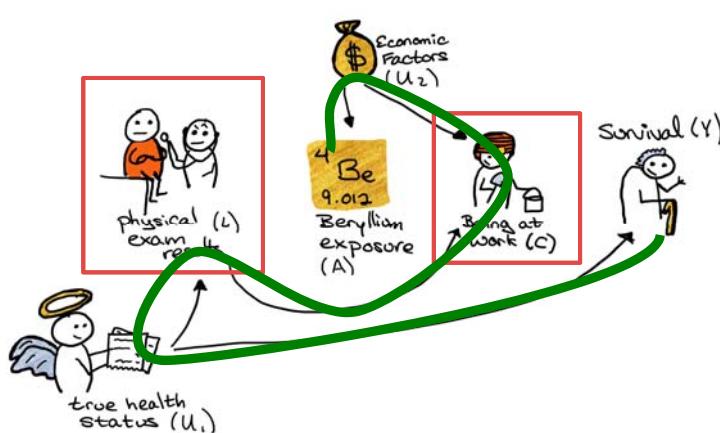
- Traditional methods to adjust for L will induce a selection bias

Alternative methods are needed to address confounding with this structure

- Ex: inverse probability weighting, g-formula



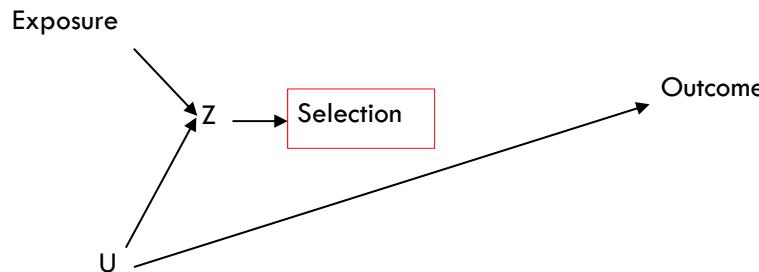
Selection Bias: Healthy worker bias



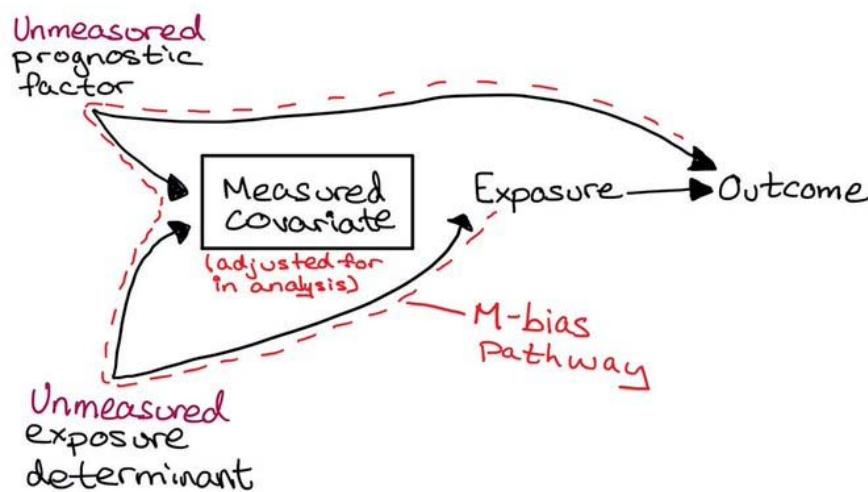
A =Beryllium exposure
 Y =Mortality
 U_1 =True health status
 C =Being at work
 L =Physical exam
 U_2 =Economic factors

Getting more careful with colliders

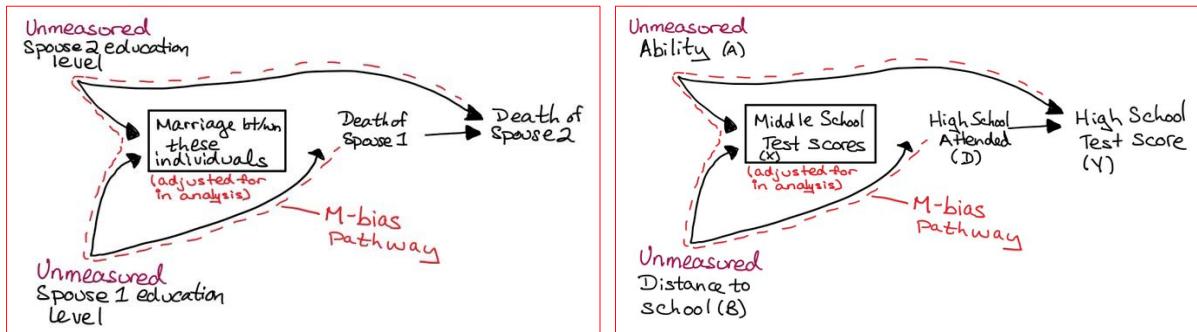
- Colliders can be opened by conditioning on them, but that's not all
- Colliders can also be opened by conditioning on their descendants



Colliders can also be pre-treatment: M-bias



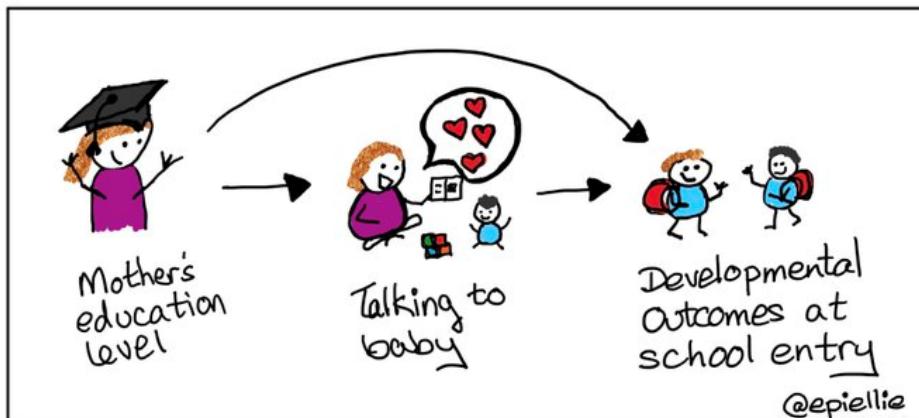
Colliders can also be pre-treatment: M-bias



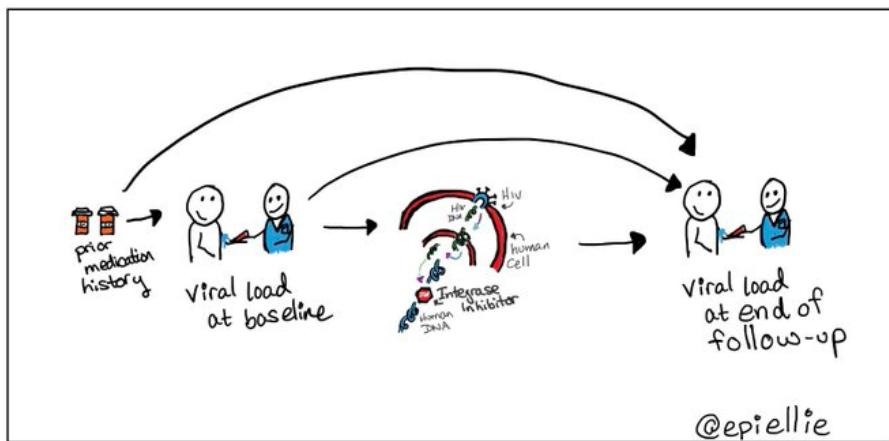
Using DAGs in applied research

1. Start with the exposure(s), outcome(s), and restriction criteria
2. Add any known & measured common cause of each pair of these, even if you're not sure of a causal relationship
3. Consider unknown variables and add placeholders
4. Check for open paths & identify how to block them (tools like DAGITTY can help)

Pop-Quiz: What do we need to condition on to estimate the effect of talking to baby on school readiness?

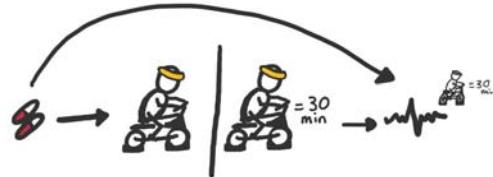


Pop-Quiz: What do we need to condition on to estimate the effect of integrase inhibitor use on viral load?



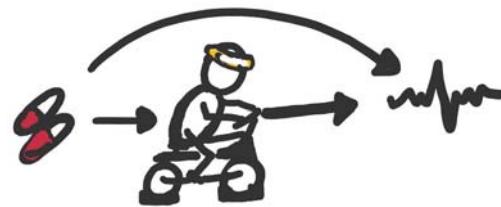
Limitations of DAGs

- Not easy to depict effect modification
- Non-parametric & don't incorporate strength of relationships
- Hard to decide when to stop building out the DAG
- Can't easily see the causal question
 - Solution to this last one is the Single World Intervention Graph (SWIG)

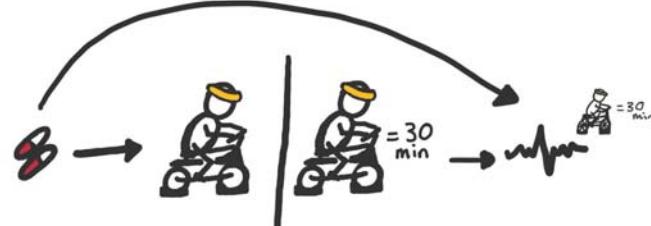


A short introduction to SWIGs

DAG: shows nodes & edges in the real world



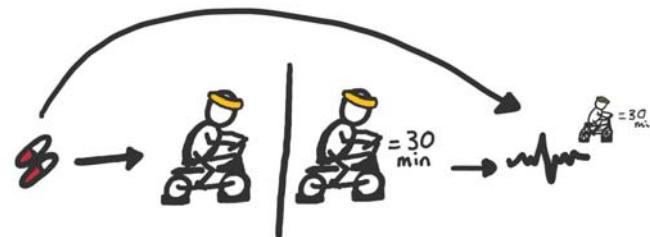
SWIG: shows nodes & edges in a world where you intervene



Why are SWIGs useful?

We can see exactly what counterfactual we are interested in on the graph

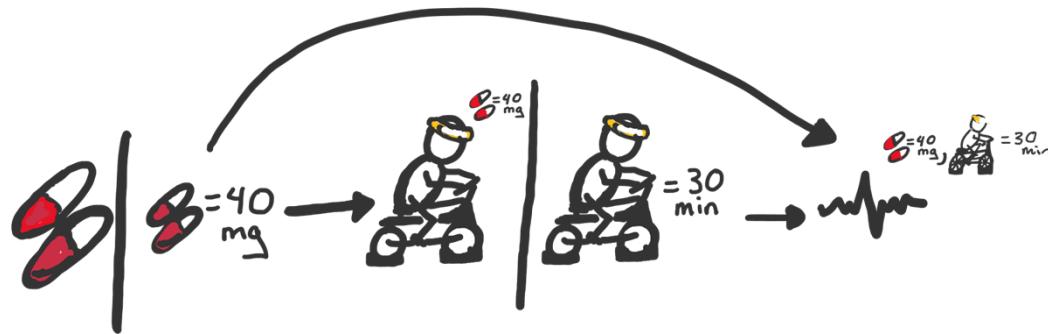
This is particularly useful when we are interested in mediation...



Mediation SWIGs

Average Blood pressure if we had intervened to set medication to 40 mg/ day and exercise to 30 min/day

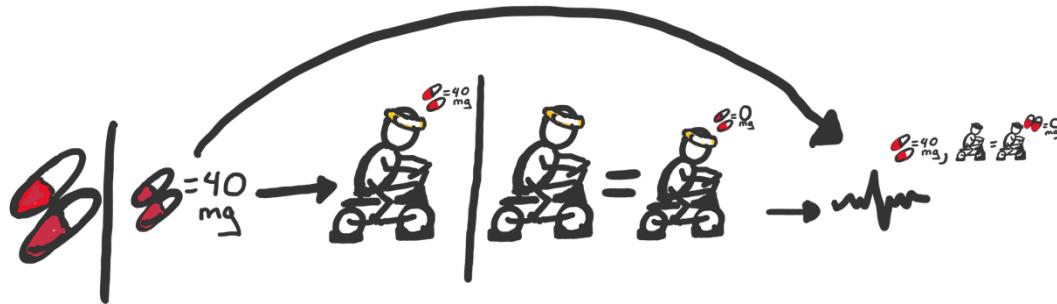
- Total effect: all arrows together
- Controlled direct effect: top arrow only



Mediation SWIGs

Natural effects: Average Blood pressure if we had intervened to set medication to 40 mg/ day and exercise to what it would have been if medication had been 0 mg / day

- Clear from the SWIG that we can't estimate these because they rely on cross-world information



Discussion time!

Have you used a DAG in your research?

- How did you build it? Systematic lit review? Expert discussion? Ad hoc lit review?
- How did you use it? Design the study? Understand the results?



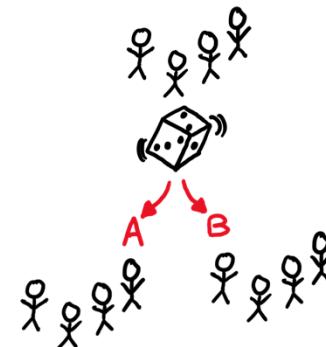
Where to get more information: Part III DAGs

- Causal Inference: What If, Hernan & Robins. Available online at:
<https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>
- Causal Diagrams for Epidemiologic Research. Greenland, Pearl, Robins. 1999. *Epidemiology*; 10(1):37-48
- Draw your Assumptions before your Conclusions. Harvard EdX Course:
<https://online-learning.harvard.edu/course/causal-diagrams-draw-your-assumptions-your-conclusions>
- Murray. Graphical models for causal inference using LaTeX. Available online at: https://github.com/eleanormurray/causalgraphs_latex

Lunch Break



Part IV: Intention-to-treat effects, but better



Reminder: in a (pragmatic) trial we can learn about

1. Intention-to-treat effects

- Effect of **randomization** to treatment

2. Per-protocol effects: effect of **treatment**

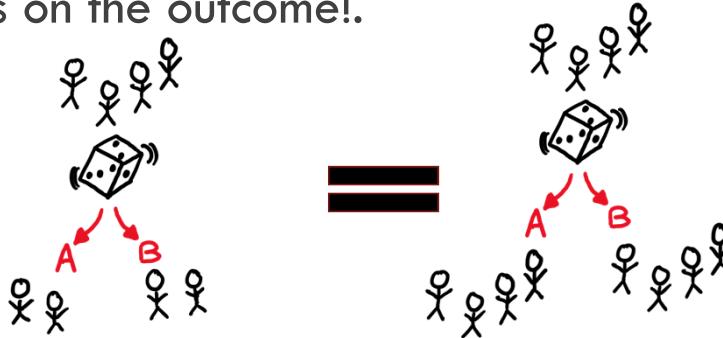
- Effect of **initiating** treatment
- Effect of **adhering** to treatment protocol
- Effect of **receiving** point intervention, **among the 'compliers'** (not necessarily all adherers!)

Effect of randomization is not really an interesting effect

- Often a **lower bound** on the effect of treatment compared to placebo
- If no blinding, includes effect of **any other actions** patients take
- Lower bound is **insufficient** for adverse events or safety
- When comparing active treatments, ITT can vary towards **or away from** the null
- No **real world**, clinical, equivalent of randomization

We **CAN** get an unbiased estimate of the intention-to-treat effect...

If we have no loss to follow-up then a simple comparison between randomization status is an unbiased estimator of the effect of randomization status on the outcome!



... but unbiased doesn't mean interesting!

If no one adheres at all, then the true ITT and our estimated ITT will both be null!

Correct but unhelpful!

If the proportion that adhere in study 1 and study 2 are different, then the estimated ITT in each study will equal the true ITT among that sample, but the estimated ITTs will differ between studies!

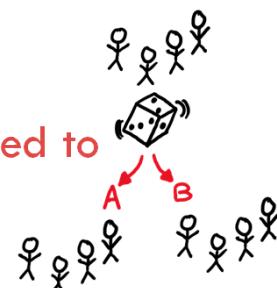
Correct but confusing!

If we have loss to follow-up, then estimating the ITT is harder!

Randomization ensures* no confounding at baseline for treatment assignment

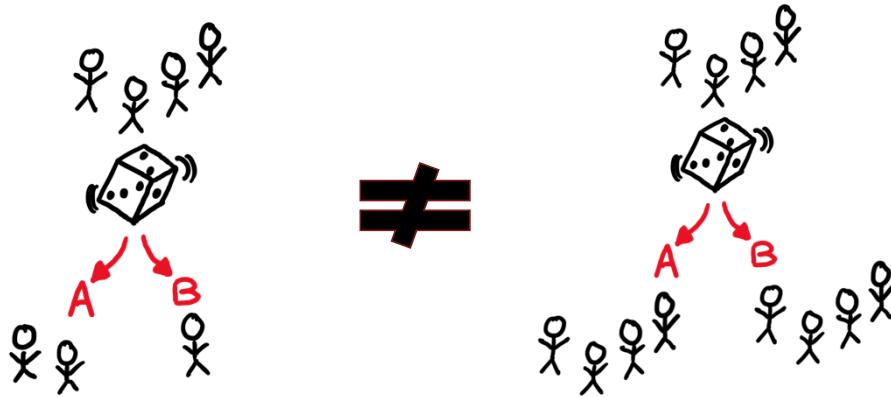
- Treatment happens after randomization
- Loss to follow-up happens after randomization

Post-randomization events are not guaranteed to be unconfounded!

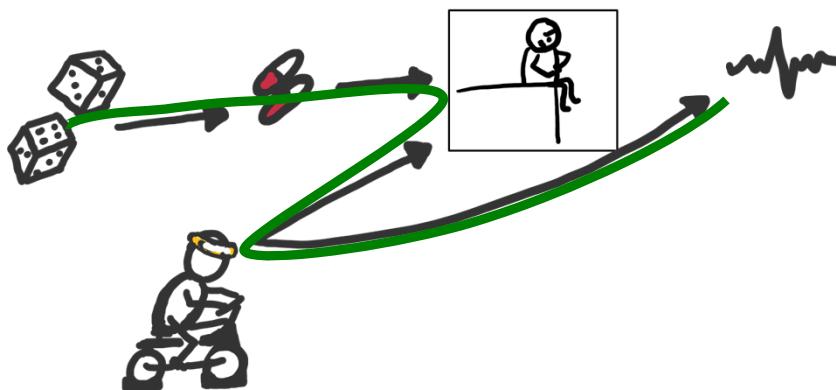


*Assuming sufficient sample size & adequate randomization

Loss to follow-up means we (probably) can't just compare outcomes



Visualizing loss to follow-up on a DAG



Activity: Selection bias is a toss-up

Hypothesis 1: when we toss two pairs of coins, the result of each coin toss is independent

Hypothesis 2: when we look only among pairs with at least one head, the proportion of heads & tails is not independent

Activity: Selection bias is a toss-up

1. Groups of 2

- 1 person is the record keeper:
<http://bit.ly/CoinsToss>

2. Flip your coin 10 times each

- Record keeper: for each pair of tosses select the appropriate combination of heads / tails check-boxes

Selection Bias Coin Toss activity				
A paired coin toss game, choose a group name and then record the results of each pair of coin tosses				
<input type="text"/> Group name <input type="text"/> Your answer				
Coin	Partner 1 - heads	Partner 1 - tails	Partner 2 - heads	Partner 2 - tails
1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Let's analyze the results!

First, comparing all pairs of coin tosses:

		Coin 2	
		Tails	Heads
Coin 1			
Tails			
Heads			

(Code is available in R on the github if you want to try it too)

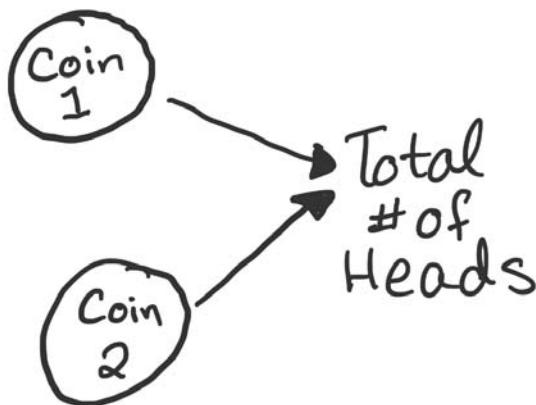
Let's analyze the results!

Next, restricting to pairs where at least 1 coin has a head!

		Coin 2	
		Tails	Heads
Coin 1			
Tails			
Heads			

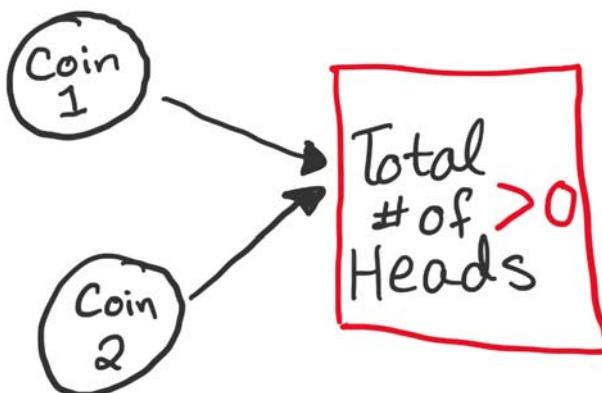
(Code is available in R on the github if you want to try it too)

Why do we see these results?



- The outcome of each coin is independent
- Both coins contribute to the total head count

Why do we see these results?



- Conditional on at least one head, the outcome of the coins are no longer independent
- We have created selection bias!

What does this mean for our trials?

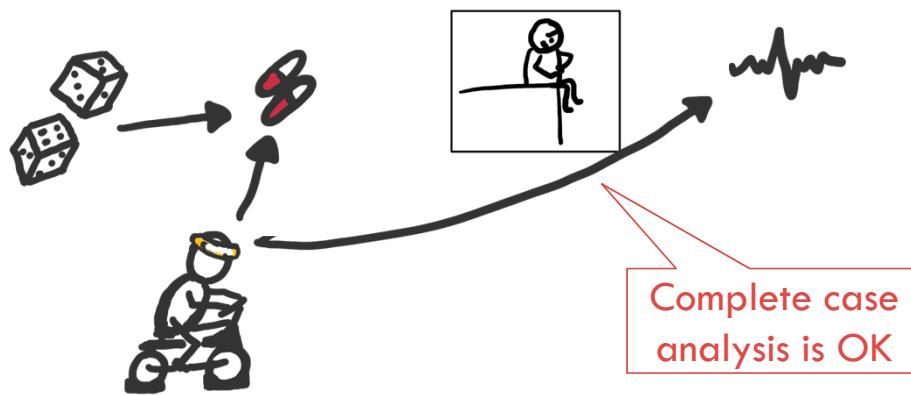
We'll almost never have no loss to follow-up, but we have options for analysis **depending on our assumptions**

Do we believe loss to follow-up is:

- complete random?
- dependent on measured covariates only?
- dependent on measured covariates AND random assignment?
- dependent on unmeasured covariates?

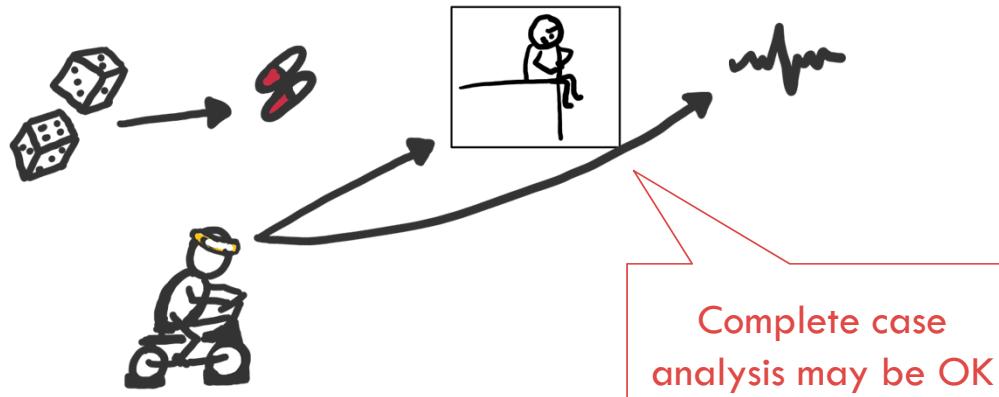
Missing completely at random (MCAR)

If we believe loss to follow-up is completely random, we are assuming the following DAG is true:



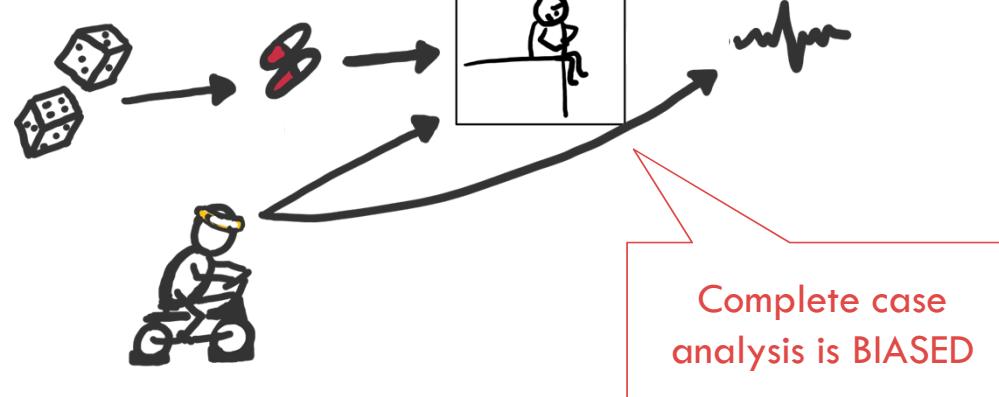
Missing at random-ish (MAR)

If we believe loss to follow-up is dependent on measured covariates only, we are assuming the following DAG is true:



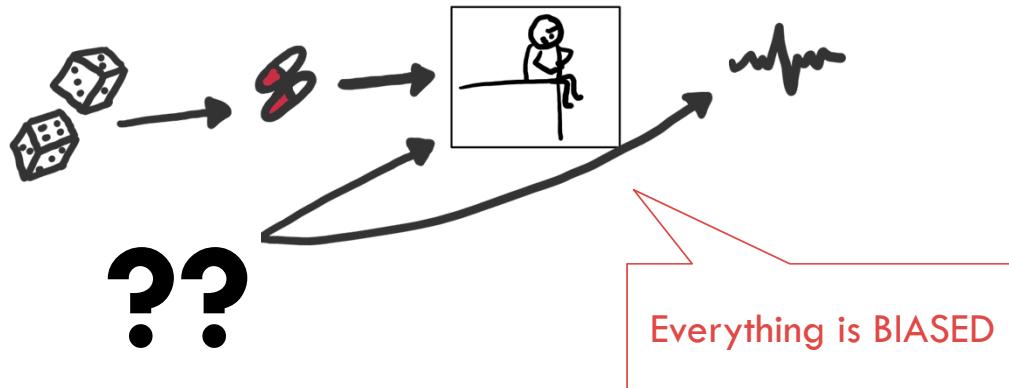
Missing at random (MAR)

If we believe loss to follow-up is dependent on measured covariates AND exposure, we are assuming the following DAG is true:



Missing not at random (MNAR)

If we believe loss to follow-up is dependent on **unmeasured covariates**, we are assuming the following DAG is true:



What to do when we have MAR loss to follow-up?

Option 1: multiple imputation – guess unknowns based on conditional distribution of knowns

- Caution: creates many large datasets that need to be combined

What to do when we have MAR loss to follow-up?

Option 2: Parametric g-formula – simulate what would have happened if no one had been lost to follow-up based on the conditional distribution of knowns

- Caution: can be computationally somewhat intense

What to do when we have MAR loss to follow-up?

Option 3: Inverse probability of censoring weights – up-weight people who complete follow-up and are similar to those who are lost to follow-up to compensate

- Caution: can have lower statistical efficiency when sample sizes are small

Let's get to know a g-method better!

1. **Inverse probability weighting** (aka IPW) of marginal structural models

2. (Parametric) **G-formula**

3. **Doubly-robust** estimation (aka targeted maximum likelihood estimation or TMLE if estimated using machine learning)

4. **G-estimation** of structural nested models



A closer look at inverse probability weights

Inverse probability weighting is a way of correcting for missing information.

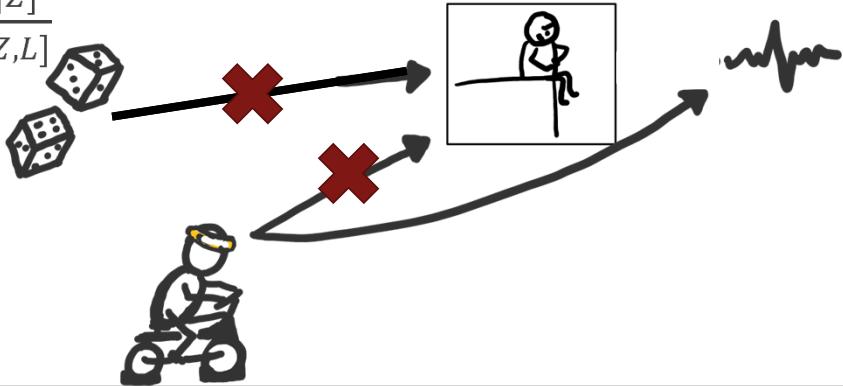
We can correct for:

- Loss to follow-up: Missing outcome under assigned treatment
- Non-adherence: Missing counterfactual outcome, had they received assigned treatment
- Other missingness:
 - E.g. when visits are missed but we wanted to look at info collected at those visits.

Inverse probability weights for a single time point

- $W = \frac{1}{\Pr[C=0|Z,L]}$

- $SW = \frac{\Pr[C=0|Z]}{\Pr[C=0|Z,L]}$



Inverse probability of weights for multiple time points

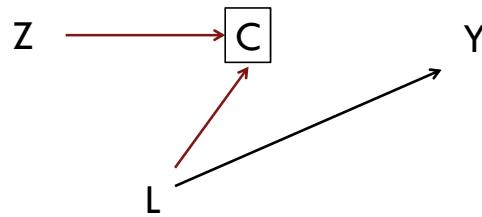
- $W_t = \prod_{j=0}^t \frac{1}{\Pr[C_j=0|Z, \bar{L}_j, \bar{C}_{j-1}=0]}$

- $SW_t = \prod_{j=0}^t \frac{\Pr[C_j=0|Z]}{\Pr[C_j=0|Z, \bar{L}_j, \bar{C}_{j-1}=0]}$

- At each time, each person receives a weight inversely proportional to the probability of remaining uncensored, conditional on randomization and time-varying covariates

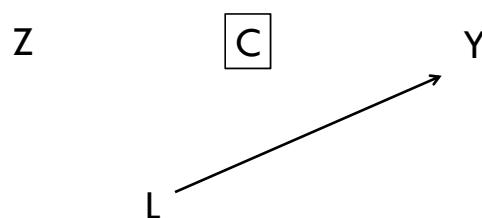
Adjusting for loss to follow-up

$$\blacksquare W_t = \prod_{j=0}^t \frac{1}{\Pr[\textcolor{red}{C}_j=0 | \textcolor{red}{Z}, \bar{\textcolor{red}{L}}_j, \bar{C}_{j-1}=0]}$$



Adjusting for loss to follow-up

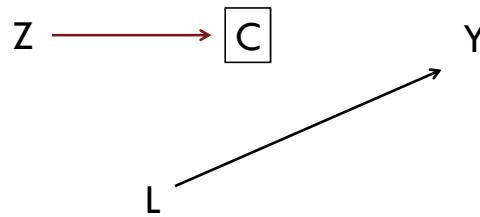
$$\blacksquare W_t = \prod_{j=0}^t \frac{1}{\Pr[\textcolor{red}{C}_j=0 | \textcolor{red}{Z}, \bar{\textcolor{red}{L}}_j, \bar{C}_{j-1}=0]}$$



Non-stabilized weights
create a pseudo-population
with no selection!

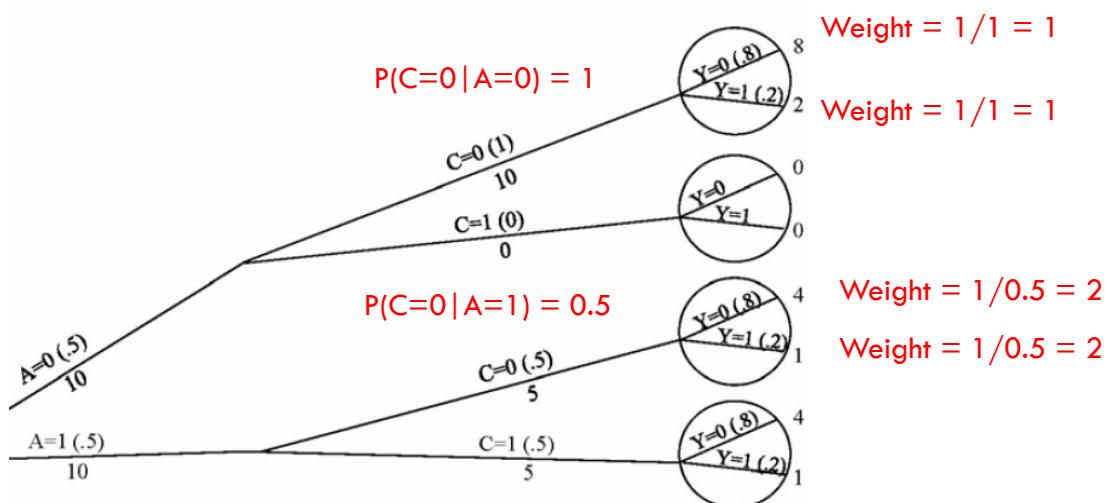
Adjusting for loss to follow-up

$$\blacksquare SW_t = \prod_{j=0}^t \frac{\Pr[C_j=0|Z, \bar{C}_{j-1}=0]}{\Pr[\textcolor{red}{C}_j=0|Z, \bar{L}_j, \bar{C}_{j-1}=0]}$$

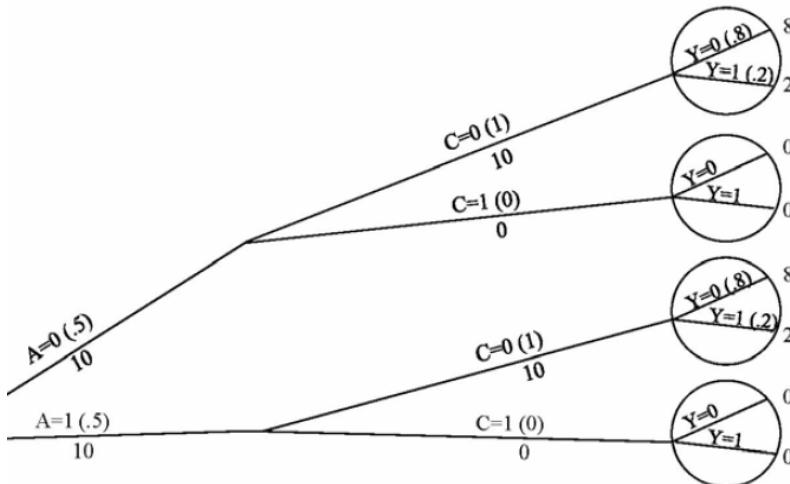


Stabilized weights create a pseudo-population
with selection BUT no selection BIAS!

Adjusting for loss to follow-up: actual sample



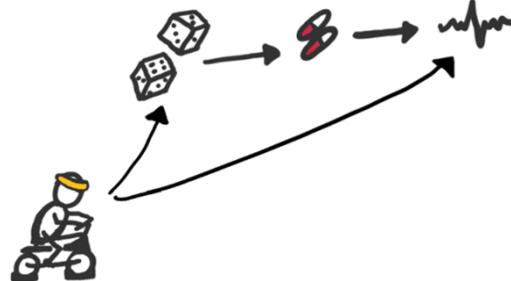
Adjusting for loss to follow-up: pseudopopulation



Other threats to the intention-to-treat effect

Random confounding:

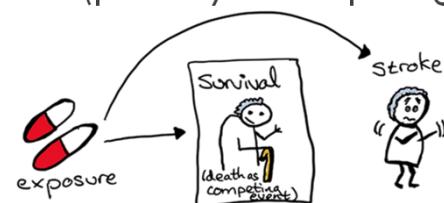
- If our trial is small, or our method of assigning treatment is not really random, we may still have baseline confounding even for the ITT
- Solution: pre-specify important prognostic factors to adjust for



Other threats to the intention-to-treat effect

Competing events:

- If our outcome can only occur among people who survive some competing event, then simply saying the “intention-to-treat” effect is insufficient.
- Assignment to A could decrease the outcome by **increasing** the competing event
- Solution: report the risk of each outcome (primary & competing)



Discussion time!

Have you ever estimated an intention-to-treat effect?

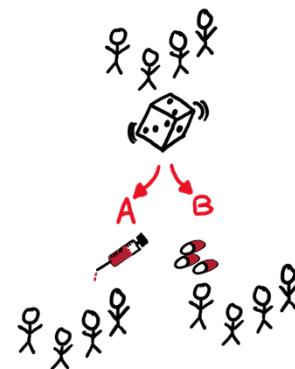
- How did you deal with loss to follow-up? What assumptions did that require?
- How did you deal with baseline covariates? Were there competing events? If so, how did you define the question?



Where to get more information: Part IV intention-to-treat

- Causal Inference: What If?, Hernan & Robins. Available online at:
<https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>
- <https://www.hsph.harvard.edu/causal/pragmatictrials/>
- Murray et al. 2018. J Clin Epi 103:10-21.
- Hernan & Scharfstein. 2018, Ann Intern Med; 168(7):515-6.

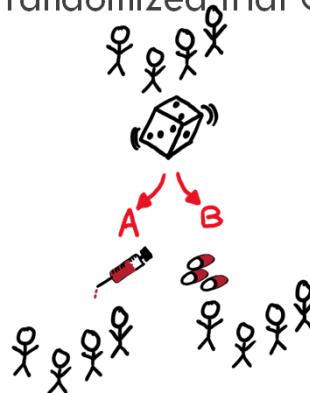
Part V: Per-protocol effects, done right



What is the per-protocol effect?

The effect of receiving treatment A versus receiving treatment B

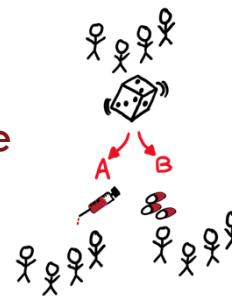
- Can be estimated in a randomized trial OR an observational study



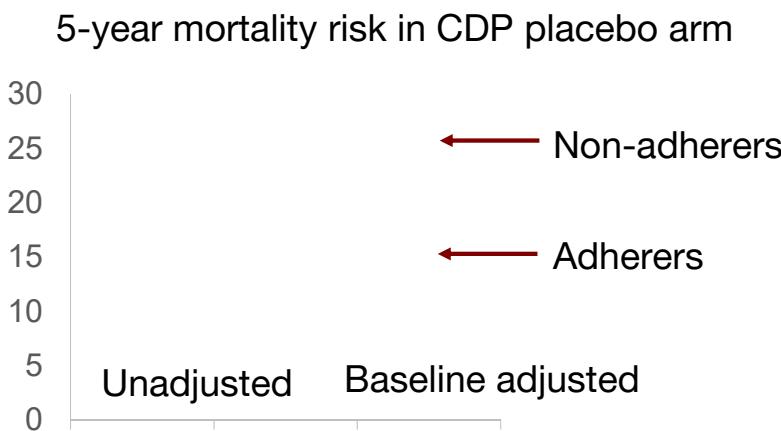
Effect of **treatment** is an interesting effect

- Relevant for **real world**, clinical, decision making
- Allows better **risk assessment** for adverse events or safety
- **Interpretable** for both placebo and active / usual care comparators

**Per-protocol effect is the
effect we really want!**

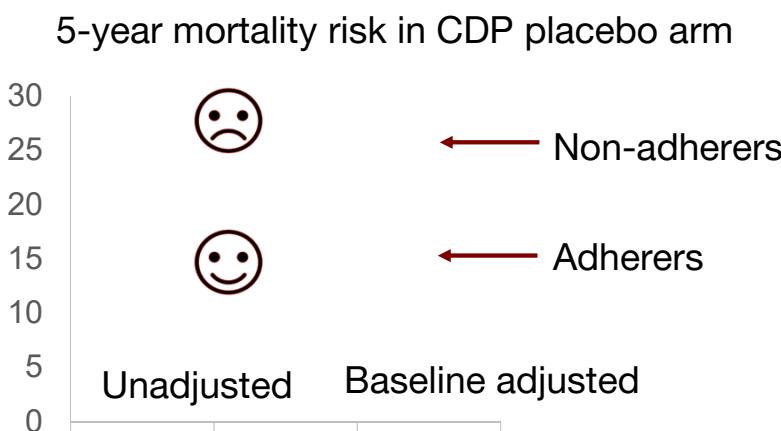


But isn't adherence *intractably* confounded?



Coronary Drug Project. 1980, NEJM; 303: 1038-41.

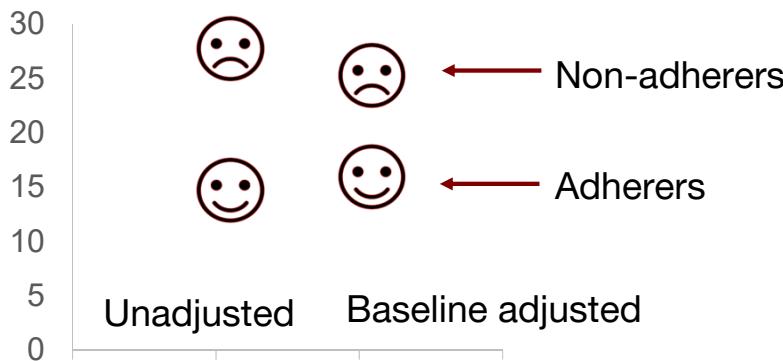
But isn't adherence *intractably* confounded?



Coronary Drug Project. 1980, NEJM; 303: 1038-41.

But isn't adherence *intractably* confounded?

5-year mortality risk in CDP placebo arm



Coronary Drug Project. 1980, NEJM; 303: 1038-41.

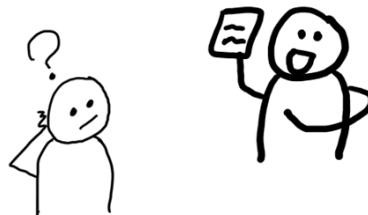
Yes, adherence is confounded...

... but it's not necessarily **intractably** so!

Confounding is an observational data problem, and we have observational data solutions we can use to fix it!

But we can't just rely on the data alone

All causal effects require that we make **assumptions** that cannot be verified in the data

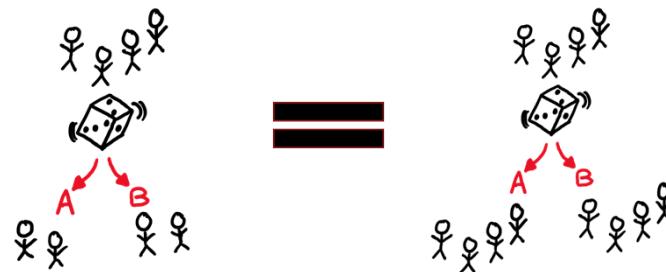


What assumptions do we need?

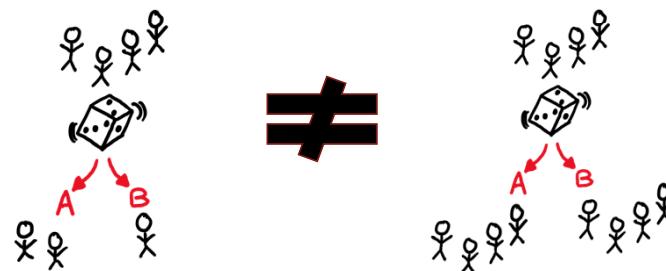
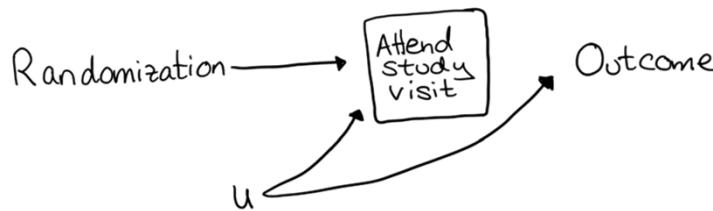
1. **No unmeasured confounding:** all common causes of the treatment and outcome are known and measured in the data
2. **No conditioning or restricting on common effects:** for trials, the important part is no loss to follow-up without appropriate adjustment.



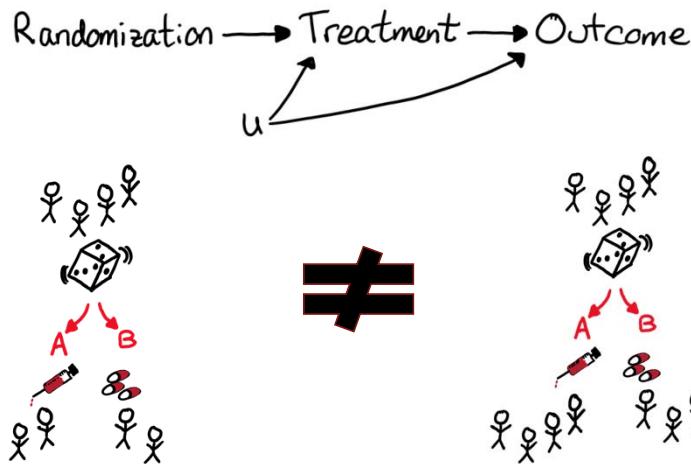
Intention-to-treat effect has no unmeasured confounding for randomization



Intention-to-treat effect might have bias from loss to follow-up



Per-protocol effect almost always requires adjustment for confounding (& loss to follow-up)

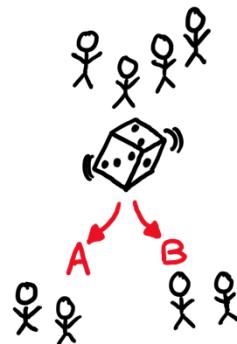


What assumptions do we need?

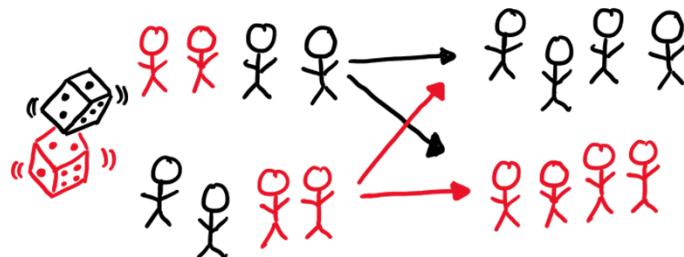
2. **Positivity:** non-zero probability of all levels of treatment for all individuals in our target population (i.e. variability in exposure)



Randomization guarantees positivity for the intention-to-treat effect



Per-protocol effects may not have positivity: consider whether violations are structural or random



Structural positivity violations require exclusion

If we cannot continue someone on treatment, then they cannot adhere to a strategy “continuously use treatment”

- Solution 1: censor as lost to follow-up at the time contraindication develops
 - Problem: per-protocol effect no longer clinically relevant
- Solution 2: change protocol to the strategy “continuously use treatment *unless* contraindication <list> develops at which time do <something else>”
 - Problem: now this is a dynamic strategy and as we'll see it's harder to analyze

Random positivity violations are okay (if uncommon)

e.g. By chance, we have no adherent 60 year olds after time $t = 13$, but we do have adherent 59 and 61 year olds still

- Solution: extrapolate from neighboring levels of the variable, just like we normally do with regression modeling

What assumptions do we need?

3. Consistency: our treatment levels are clear and well-defined

- Are we asking a specific enough question to get an answer we can understand?



Random assignment is a well-defined intervention

The intention-to-treat effect by definition has consistency because we know exactly how we are generating our random assignment

- Caveat: if we have loss to follow-up, we may no longer have consistency
- e.g. if we try to estimate what would have happened had there been no loss to follow-up → what is our well-defined intervention to achieve no losses?

How do we use these to estimate the per-protocol effect?

Positivity:

- We can check for violations of positivity in our data

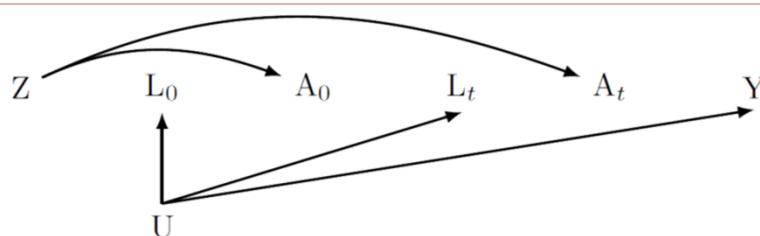
Consistency:

- We can support consistency by specifying a clear and precise protocol for adherence

No unmeasured confounding:

- We need to use expert knowledge to know what to adjust for & how to adjust for it

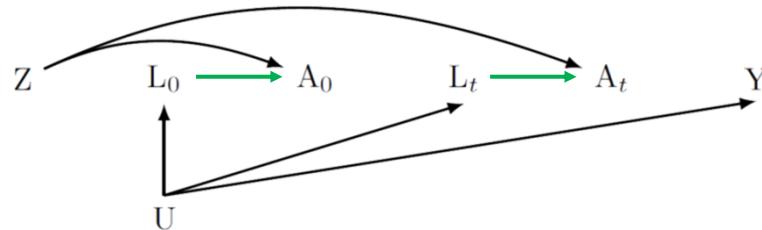
What do we need to **adjust** for?



A. Random non-adherence

- No confounding adjustment needed

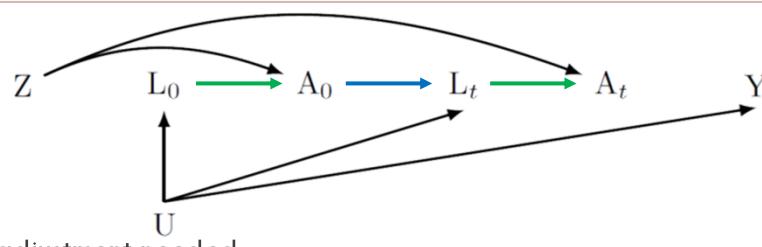
What do we need to **adjust** for?



- A. Random: No adjustment needed
- B. Adherence confounding by measured covariates**
 - Adjustment required using any method

Hernan & Robins. 2017, NEJM 377:14

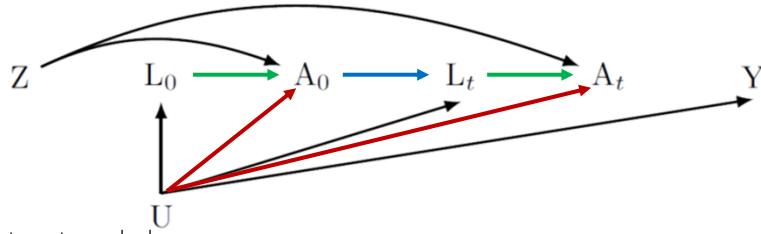
What do we need to **adjust** for?



- A. Random: No adjustment needed
- B. Measured covariates: adjustment using any method**
- C. Adherence confounding by measured covariates and prior adherence**
 - G-methods required

Hernan & Robins. 2017, NEJM 377:14

What do we need to **adjust** for?



- A. Random: No adjustment needed
- B. Measured covariates: adjustment using any method
- C. Measured covariates & adherence: g-methods
- D. Adherence confounding by measured covariates, prior adherence, and unmeasured covariates**
 - Strong assumptions + structural nested models

Hernan & Robins. 2017, NEJM 377:14

Let's get to know a g-method better!

1. **Inverse probability weighting** (aka IPW) of marginal structural models
2. (Parametric) **G-formula**
3. **Doubly-robust** estimation (aka targeted maximum likelihood estimation or TMLE if estimated using machine learning)
4. **G-estimation** of structural nested models



What is the parametric g-formula?

- A generalization (**g**) of standardization to time-varying settings
- An equation (**formula**) that relates the observational data to the counterfactual data
- Solved using Monte-Carlo simulation, which relies on (**parametric**) modeling assumptions

The general **formula** for the parametric g-formula

For a single time point of exposure:

$$\Pr[Y^a = 1] =$$

- The probability of the counterfactual outcome (**Y^a**) if everyone received exposure level $A=a$

The general formula for the parametric g-formula

For a single time point of exposure:

$$\Pr[Y^a = 1] = \sum_{l,z} P[Y = 1 | A = a, L = l, Z = z]$$

- The probability of the counterfactual outcome (Y^a) if everyone received exposure level $A=a$
- the average of stratum-specific observed outcome probabilities among people who received exposure a ($\Pr(Y=1 | A=a, L=l, Z=z)$)

The general formula for the parametric g-formula

For a single time point of exposure:

$$\Pr[Y^a = 1] = \sum_{l,z} P[Y = 1 | A = a, L = l, Z = z] P[L = l | Z = z]$$

- The probability of the counterfactual outcome (Y^a) if everyone received exposure level $A=a$
- the average of stratum-specific observed outcome probabilities among people who received exposure a ($\Pr(Y=1 | A=a, L=l, Z=z)$)
- weighted by the probability of being in each stratum ($\Pr[L=l, Z=z]$)

The general formula for the parametric g-formula

For a single time point:

$$\Pr[Y^a = 1] = \sum_l \Pr[Y = 1 | A = a, L = l] \Pr[L = l]$$

For multiple time points:

$$\begin{aligned} \Pr[Y_{k+1}^g = 1] &= \sum_{\bar{l}_k} \sum_{j=0}^k \Pr[Y_{j+1} = 1 | \bar{L}_j = \bar{l}_j, \bar{A}_j = \bar{a}_j^g, \bar{Y}_j = 0] \\ &\quad \times \prod_{s=0}^j [\Pr[Y_s = 0 | \bar{L}_{s-1} = \bar{l}_{s-1}, \bar{A}_{s-1} = \bar{a}_{s-1}^g, \bar{Y}_{s-1} = 0] \\ &\quad \quad \quad \times f(l_s | \bar{l}_{s-1}, \bar{a}_{s-1}^g, \bar{Y}_{s-1} = 0)] \end{aligned}$$

A note about IPW for adherence

$$W^c = \frac{1}{\Pr[C=0|Z,L]}$$

Probability of NOT being censored

$$SW^c = \frac{\Pr[C=0|Z]}{\Pr[C=0|Z,L]}$$

$$W^a = \frac{1}{\Pr[A=a|Z,L]}$$

Probability of having each treatment level

$$SW^a = \frac{\Pr[A=a|Z]}{\Pr[A=a|Z,L]}$$

Revisiting the Coronary Drug Project

Replication

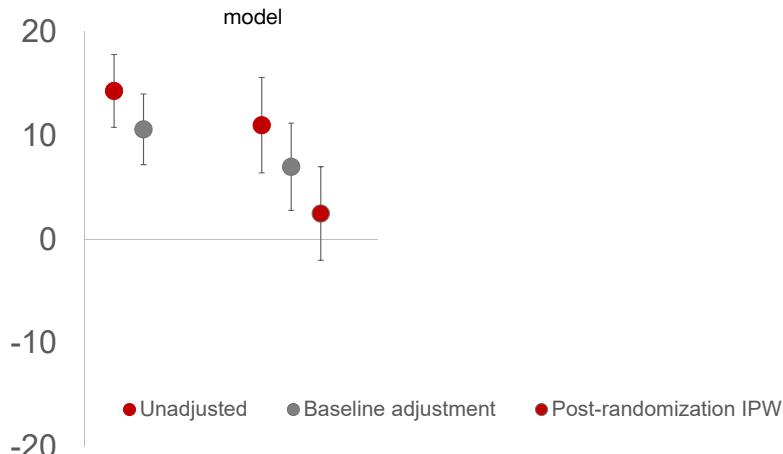


Murray & Hernan. 2016, Clin Trials 13(4): 372-8.
Murray & Hernan. 2018, Trials 19: 158.

Revisiting the Coronary Drug Project

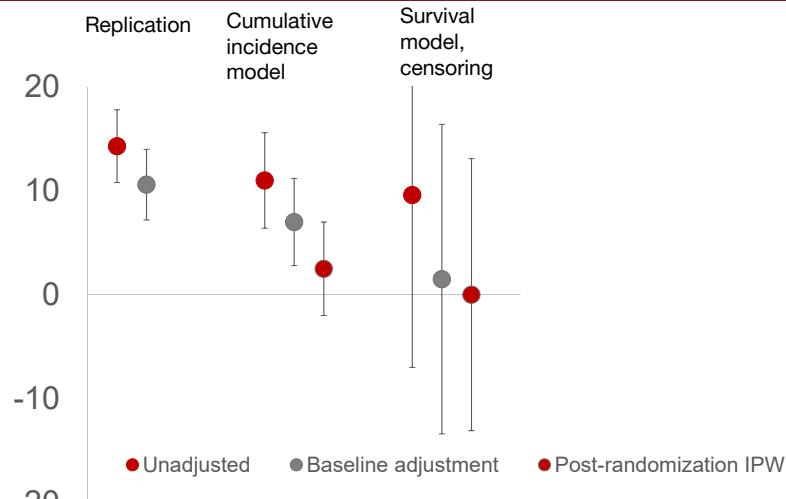
Replication

Cumulative
incidence
model



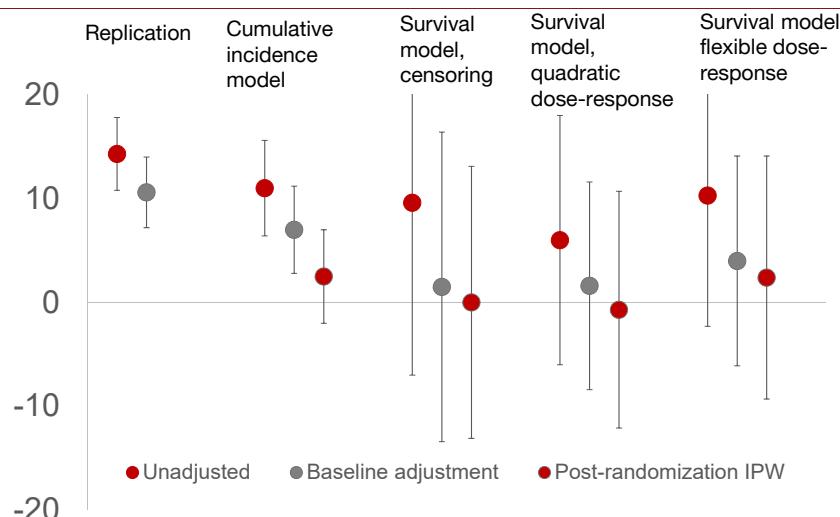
Murray & Hernan. 2016, Clin Trials 13(4): 372-8.
Murray & Hernan. 2018, Trials 19: 158.

Revisiting the Coronary Drug Project



Murray & Hernan. 2016, Clin Trials 13(4): 372-8.
 Murray & Hernan. 2018, Trials 19: 158.

Revisiting the Coronary Drug Project



Murray & Hernan. 2016, Clin Trials 13(4): 372-8.
 Murray & Hernan. 2018, Trials 19: 158.

Discussion time!

Has anyone ever estimated a per-protocol effect?

- (You probably have because, remember, this is also the causal effect we estimate in observational studies!)
- How did you choose the confounders of adherence? Should we worry about adherence causes that aren't outcome causes?



Where to get more information: per-protocol effect

- <https://www.hsph.harvard.edu/causal/pragmatictrials/>
- Murray et al. 2018. J Clin Epi 103:10-21.
- Hernan & Scharfstein. 2018, Ann Intern Med; 168(7):515-6.
- Hernan & Robins. 2017, NEJM 377:14; Lodi et al, 2016. AIDS; 30(17):2659-63.
- Murray & Hernan. 2018, Trials 19:158; Murray & Hernan. 2016, Clin Trials 13(4): 372-8.
- Causal Inference: What If?, Hernan & Robins. Available online at: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>

Afternoon Break



Guidelines for causal inference from pragmatic trials



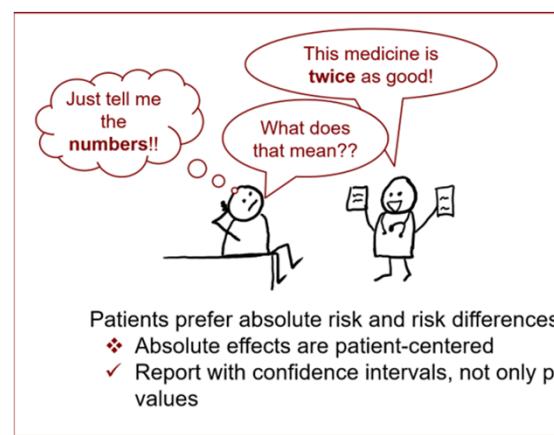
Choice of causal effect

1. To adequately guide decision making by all stakeholders, report estimates of **both** the intention-to-treat effect and the per-protocol effect, as well as methods and key conditions underlying the estimation procedure.



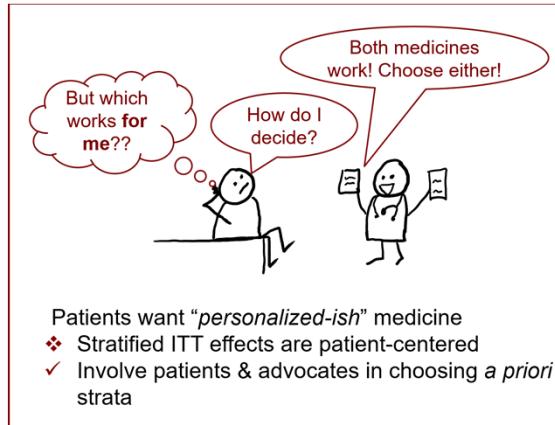
Choice of causal effect

2. Report **absolute risks** and their differences, as well as their ratios, for discrete outcomes.



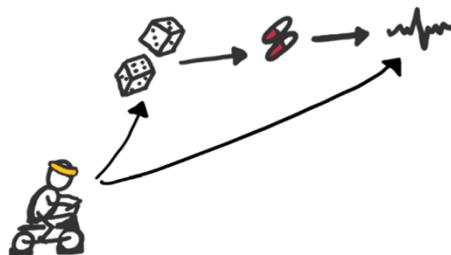
Choice of causal effect

3. Heterogeneity of treatment effects can be reported using **subgroup analyses** that use the additive scale. Patients & advocates should be included in *a priori* specification of subgroups.



Estimation of the intention-to-treat effect*

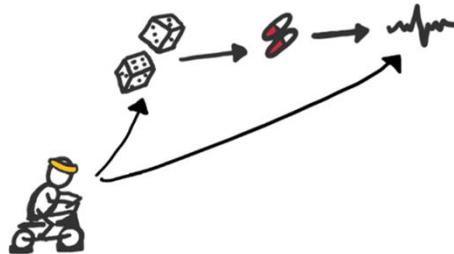
4. **Pre-specify** important prognostic factors for the outcome & the maximum acceptable difference in the distribution of these between groups; **adjust** via standardization, inverse probability weighting, or doubly-robust methods when threshold met.



* Note these guidelines also apply to per-protocol effects

Estimation of the intention-to-treat effect*

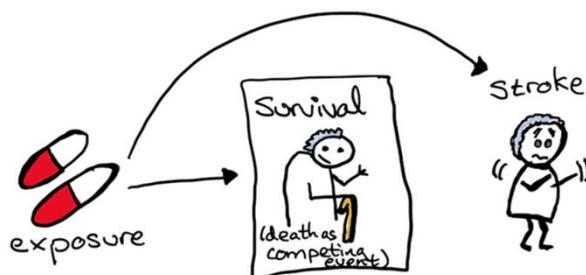
5. In **sensitivity analyses**, adjust for large imbalances in any important prognostic factors, regardless of pre-specification.



* Note these guidelines also apply to per-protocol effects

Estimation of the intention-to-treat effect*

6. In survival analyses with **competing events**, report both the risk of the competing event by treatment group and the risk of the event of interest among those who survived the competing event by treatment group.



* Note these guidelines also apply to per-protocol effects

Estimation of the intention-to-treat effect*

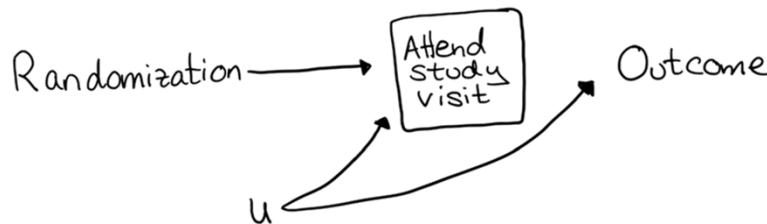
7. In survival analyses with competing events, specify the intention-to-treat effect as the total effect of treatment assignment on the outcome of interest (the simplest analysis), and justify interest in any additional effects that are estimated.



* Note these guidelines also apply to per-protocol effects

Estimation of the intention-to-treat effect*

8. Ensure that the trial protocol specifies the collection of post-randomization time-varying prognostic factors that predict loss to follow-up, and appropriately adjust for these factors to reduce selection bias.



* Note these guidelines also apply to per-protocol effects

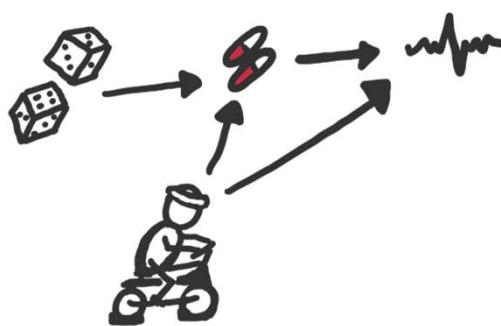
Per-protocol effects: Defining exposures strategies

Point exposures are exposures that happen at a single point in time, typically at baseline. Randomization is (usually) a point exposure.

Sustained exposures are exposures that happen repeatedly over time. Many randomized trials assign participants to sustained exposures.

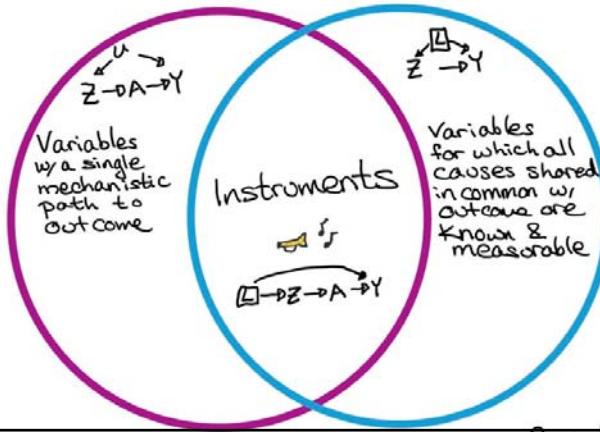
Estimating the per-protocol effect of a **point** intervention

9. Per-protocol effects of point interventions can be estimated when **sufficient data on baseline confounders** exist. Use inverse probability weighting, standardization, doubly-robust estimation, or other methods.



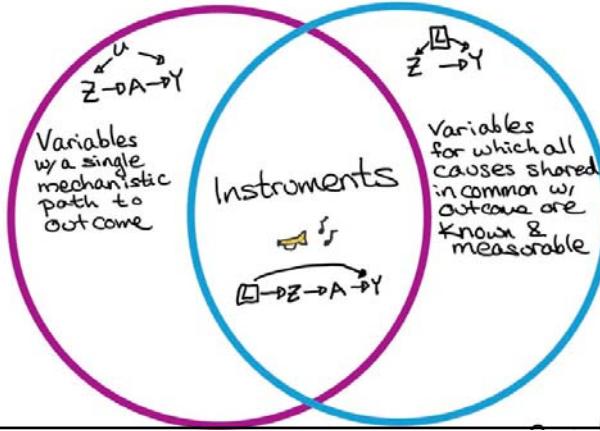
Estimating the per-protocol effect of a **point** intervention

10. Estimate **bounds** for the per-protocol effect of point interventions when the instrumental conditions are expected to hold for treatment assignment.
Provide a justification for the exclusion restriction.



Estimating the per-protocol effect of a **point** intervention

11. When instrumental conditions and monotonicity hold, discuss whether the effect in the “compliers” is of interest. If so, estimate it & provide information on the relative size and characteristics of the “compliers” subset.



Defining sustained exposures strategies

Static sustained exposures are sustained exposures that don't change over time. An example is "always eat vegetables".

Dynamic sustained exposures are sustained exposures which change over time, especially if the change is dependent on time-varying individual characteristics. An example is "take treatment unless a contraindication develops".

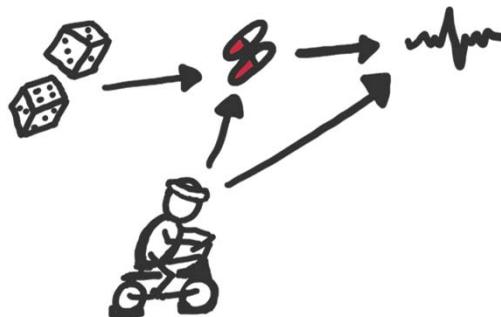
Estimating the per-protocol effect for a **sustained** intervention

12. When estimating per-protocol effects of sustained treatment strategies, specify *a priori* a **treatment protocol** that incorporates real world clinical decision-making. When there is sufficient ambiguity about appropriate strategies, more than 1 can be specified.



Estimating the per-protocol effect for a **sustained** intervention

13. Ensure **sufficient data** are collected to determine adherence throughout the follow-up, and to adjust for time-varying prognostic factors that predict adherence.



Estimating the per-protocol effect for a **sustained** intervention

14. Use **g-methods** to adjust for time-varying confounders when there is treatment-confounder feedback.

Inverse probability weighting (aka IPW) of marginal structural models

(Parametric) **G-formula**



Doubly-robust estimation (aka targeted maximum likelihood estimation or TMLE if estimated using machine learning)

G-estimation of structural nested models

Discussion time!

How do you see these guidelines applying in your work? Will it change how you design your next study?



Where to get more information

Proposed guidelines:

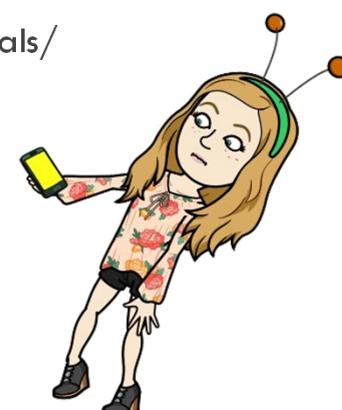
<https://www.hsph.harvard.edu/causal/pragmatictrials/>

Contact me:

 @EpiEllie

 ejmurray@bu.edu

 <https://github.com/eleanormurray>





Tomorrow: Coding Practicum

Reminder, bring a laptop with R, SAS, or Stata

Download the materials at:

https://github.com/eleanormurray/NeuRA_Sydney_2019