NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

Causal Inference in Pragmatic Trials

Eleanor Murray, PhD
Assistant Professor of Public Health
Boston University School of Public Health

Housekeeping

- All participants will be muted
- Enter **all questions** in the Zoom **Q&A/chat box** and send to Everyone
- Moderator will review questions from chat box and ask them at the end
- Want to continue the discussion? Associated podcast released about 2 weeks after Grand Rounds
- Visit impactcollaboratory.org
- Follow us on Twitter & LinkedIN:

@IMPACTcollab1 https://www.linkedin.com/company/65346172

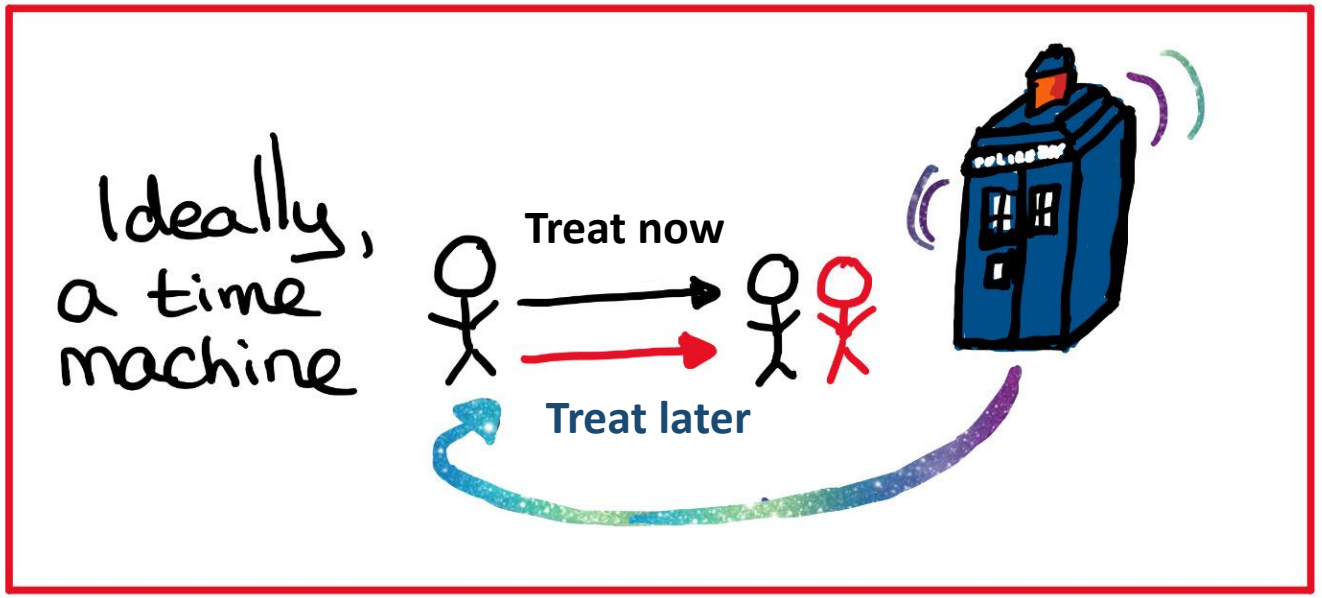
Learning Objectives

Upon completion of this presentation, you should be able to:

- Discuss the properties of a good causal research question
- Identify two key sources of potential bias common in pragmatic trials
- Recognize pragmatic trial scenarios requiring sophisticated causal inference statistical methods

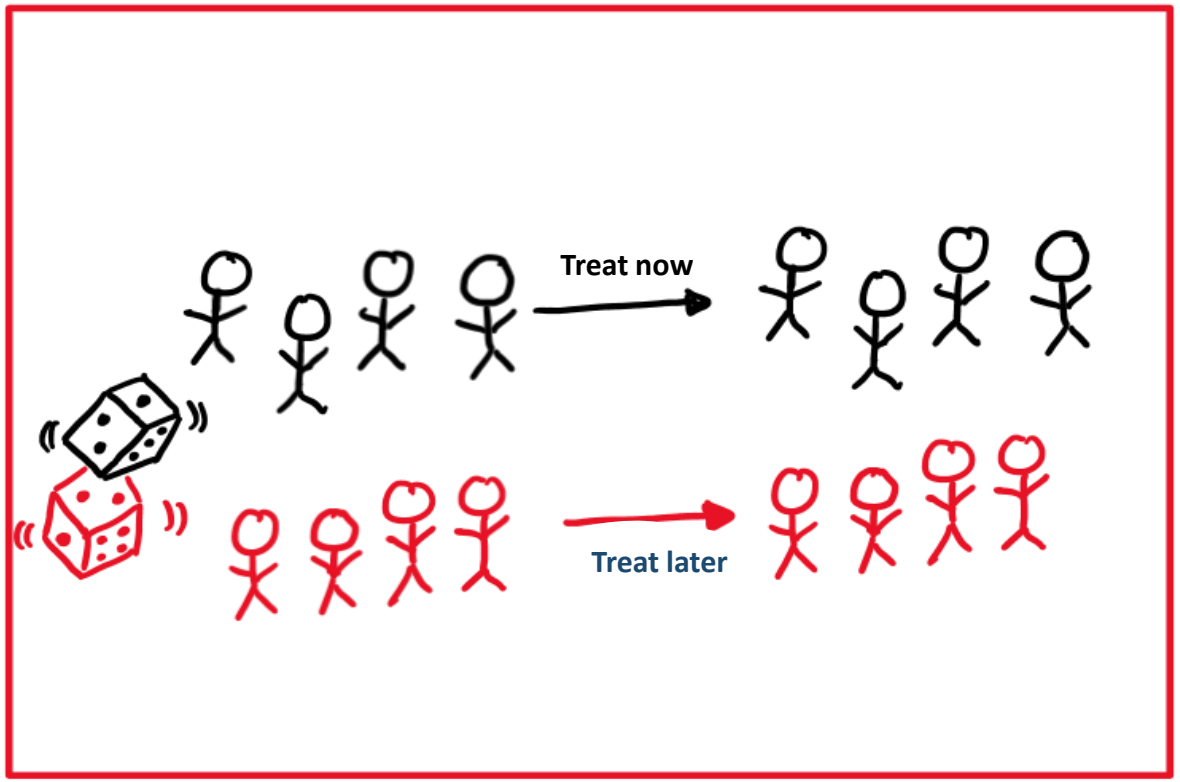
What do we want to know when we ask about causal effects?

- How would things have changed if the world had been slightly different?

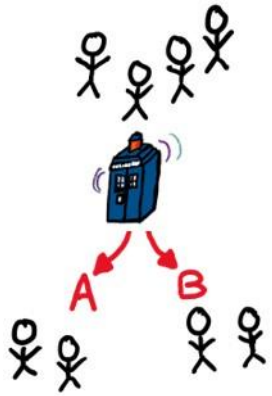


What do we want to know when we ask about causal effects?

- If we can't have a time machine, we'd like to have an ideal, perfect randomized trial.



An ideal, perfect randomized trial



NO COMPETING EVENTS

COMPLETE FOLLOW-UP

RANDOMIZATION WORKS PERFECTLY

TRIAL INTERVENTIONS MATCH CLINICAL PRACTICE

TARGET POPULATION ALL ENROLLED

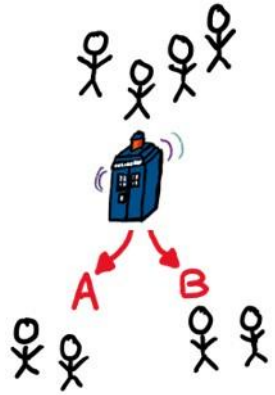
NO MEASUREMENT ERROR

NO EFFECT MODIFICATION

LARGE SAMPLE SIZE

Trials are hard, and perfect trials are impossible

- Even when we can intervene, we can't always do it perfectly!
- But often, we can't intervene at all
  - Harmful exposures
  - Time constraints
  - Cost constraints
  - Hard to access population



Tightly controlled RCTs don't always tell us what we want to know

- Randomized clinical trials are traditionally seen as optimal for detecting treatment effects…
  - randomization prevents confounding
  - but, only for randomization status not for treatment received
- But, the results can have limited applicability for clinical decision-making outside the trial
  - 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

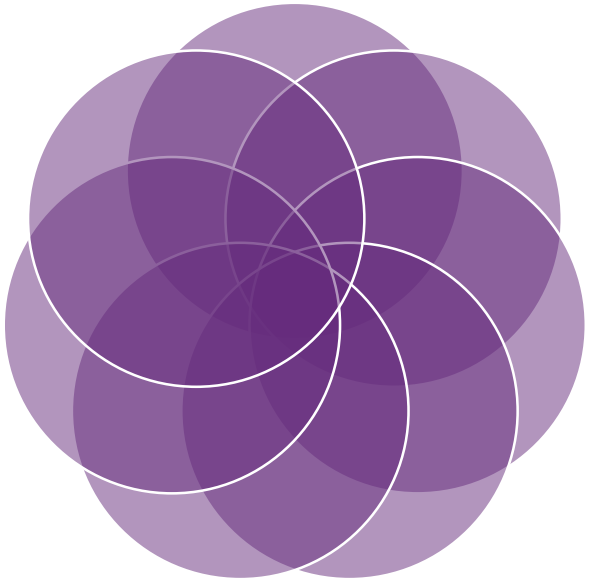
- Pragmatic randomized trials are designed to ask clinically relevant, generalizable, questions
  - Useful tools for patient-centered outcomes research
  - But, trade-off between clinical relevance and ease of use

Pragmatic trials can still have biases – (and so can explanatory RCTs)!

Informative censoring

Immortal time bias

Non-random exposure

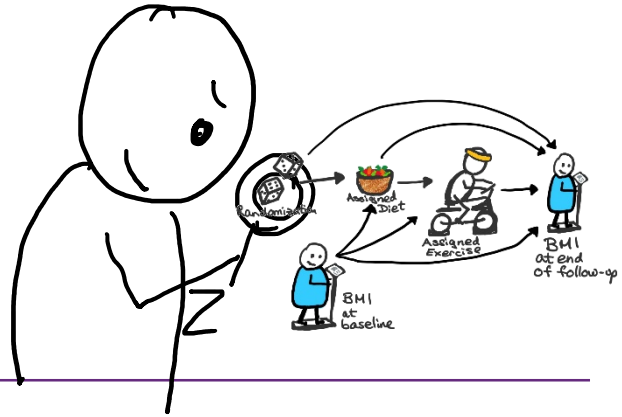


Incorrect interpretations of results

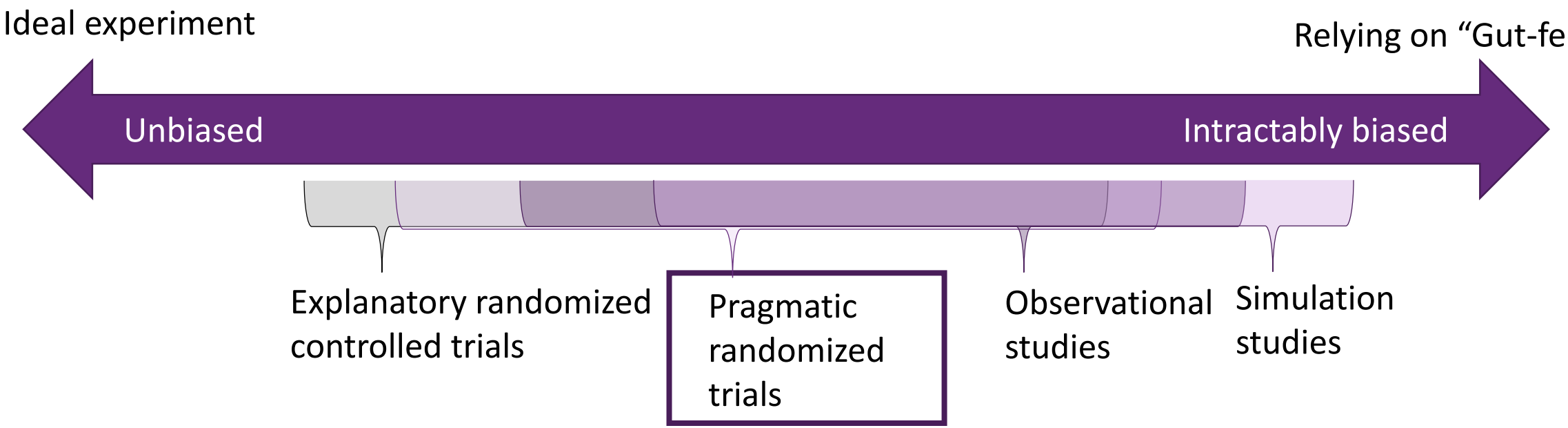
Competing events

Generalizability & transportability problems

Poorly defined causal questions



A bias continuum



Lack of generalizability

Loss to follow-up

Non-adherence

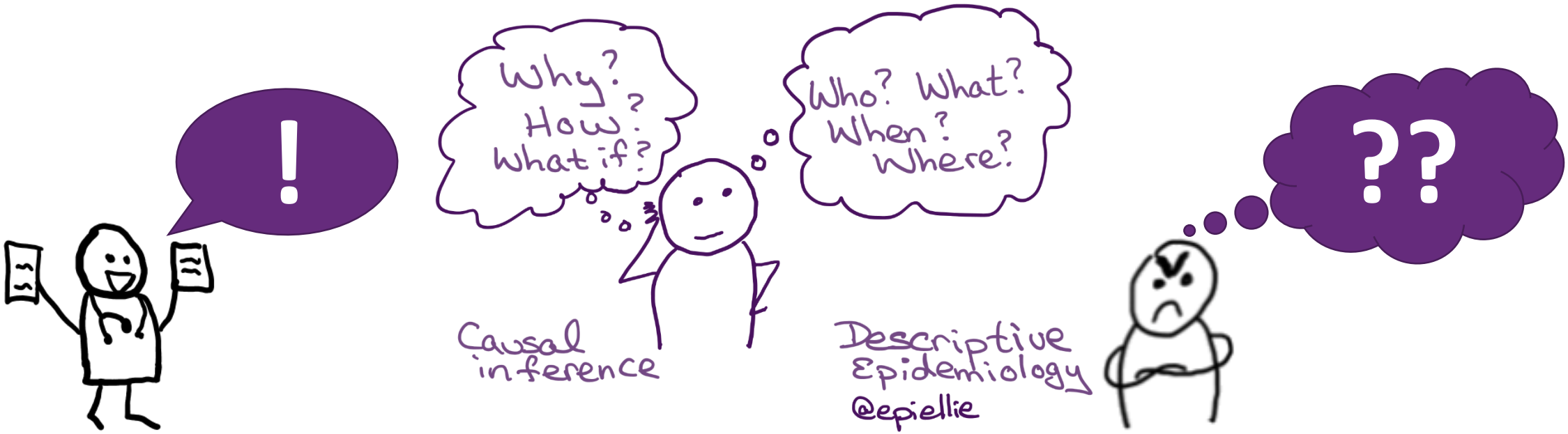
Baseline confounding

Ill-defined uncertainty

So, how do we do causal inference?

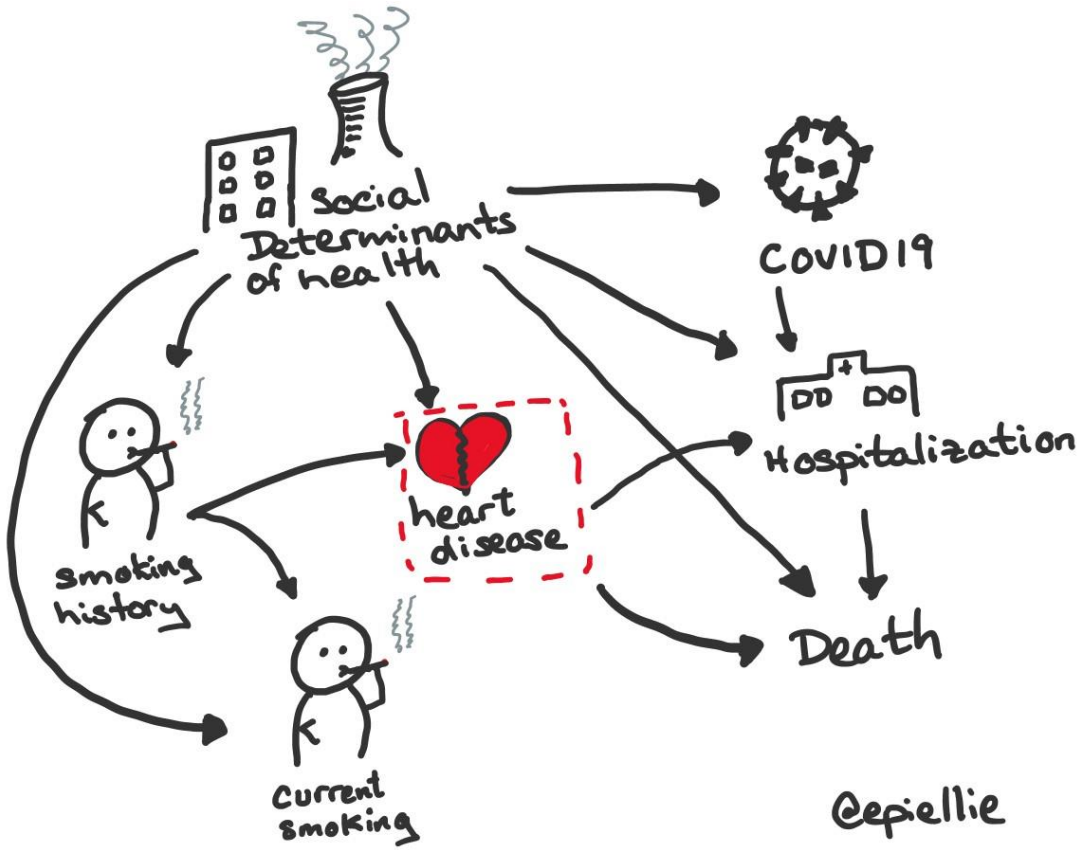
A two-step causal algorithm:

1. Ask good questions
2. Answer them with appropriate methods



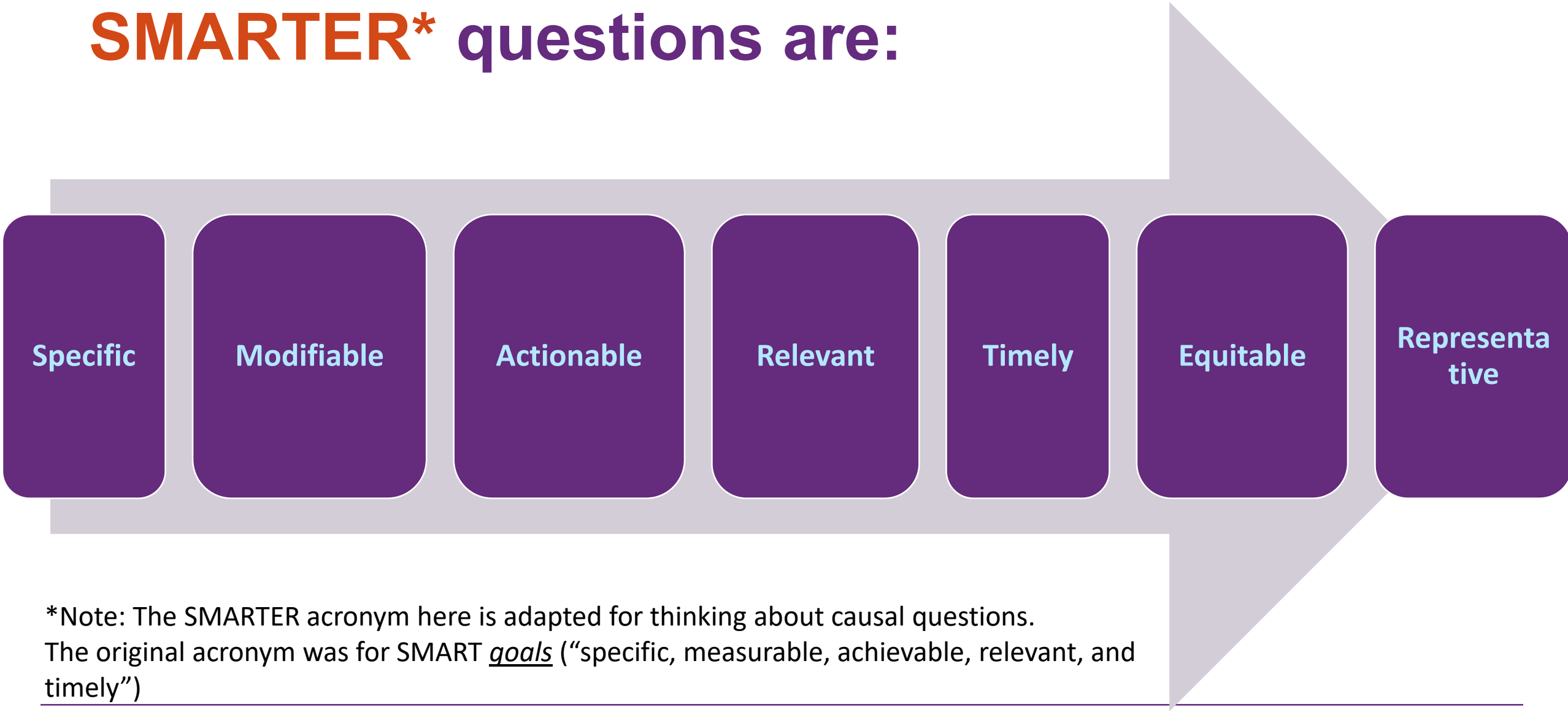
Asking good questions is hard

Health and illness have so many interacting causes



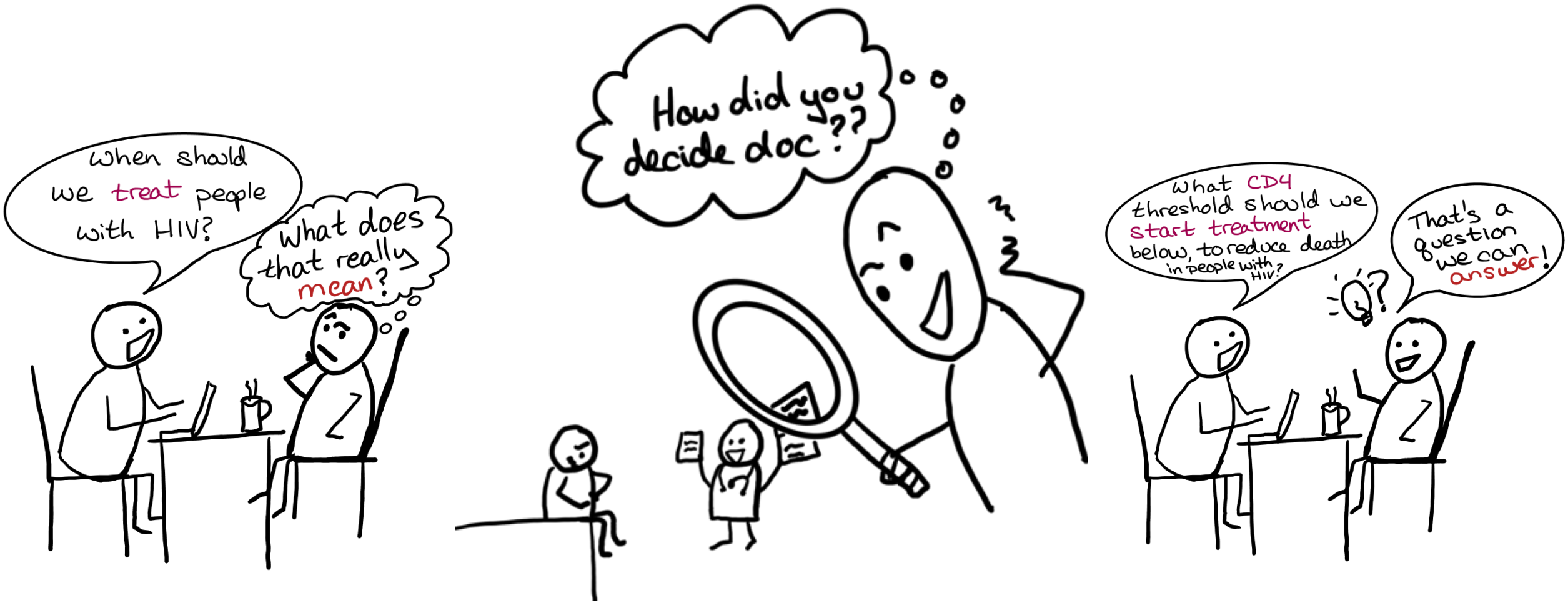
Exposures have so many different consequences

SMARTER* questions are:



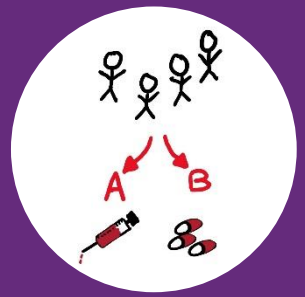
*Note: The SMARTER acronym here is adapted for thinking about causal questions.
The original acronym was for SMART *goals* (“specific, measurable, achievable, relevant, and timely”)

Rule 1 – be specific!



Specific questions inform specific decisions: by a clinician, a policy maker, a patient, etc.

Who will be making the decision, and whom are they making it for?



What will they be deciding between?

When will they be deciding and what information do they have at that time?

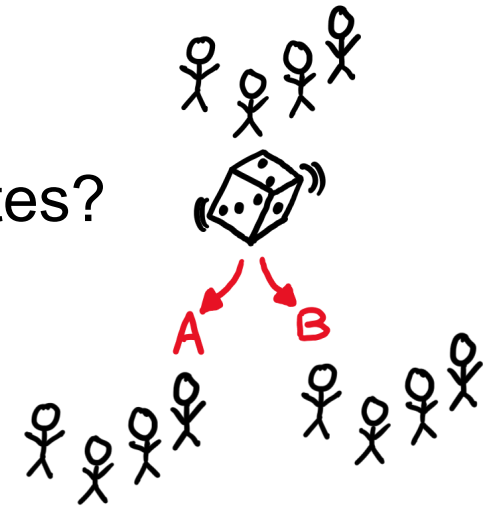
Where is the decision happening?

Why is the decision being made?

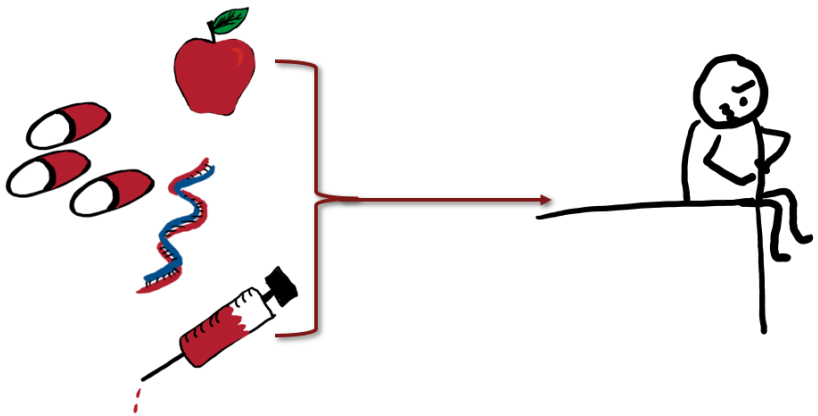
NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

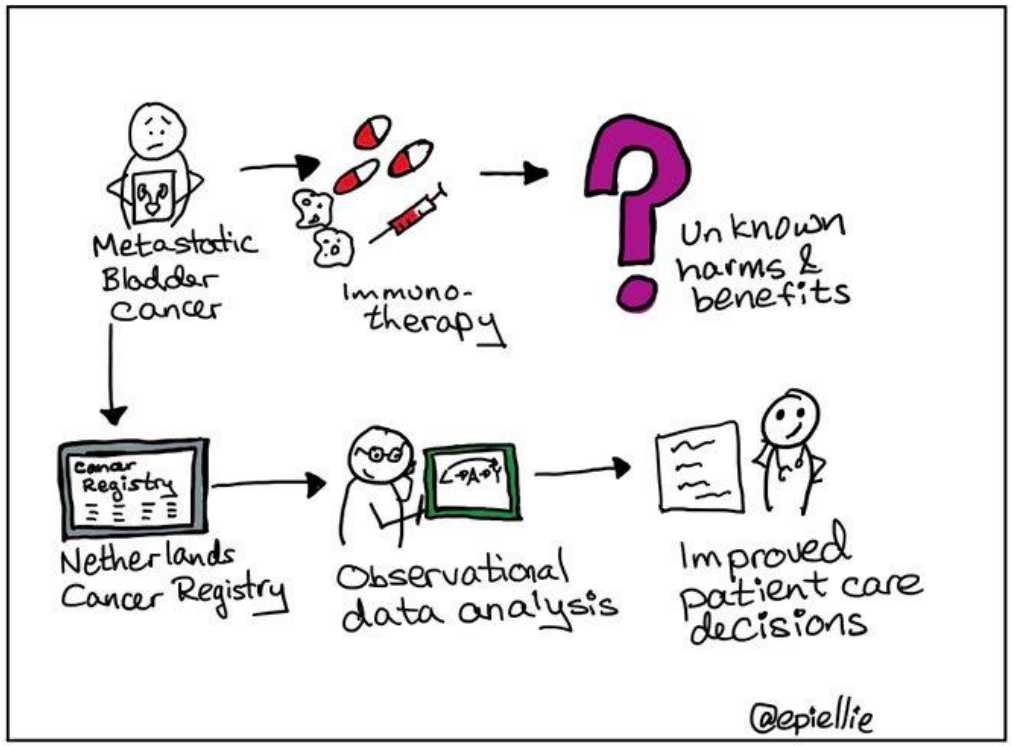
Pragmatic trials and specific questions

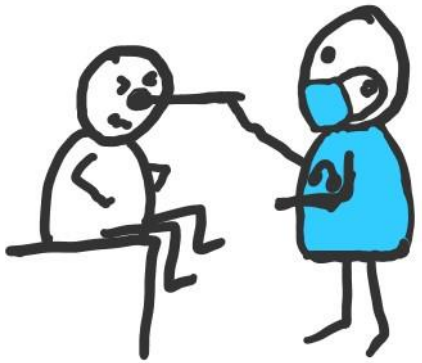
- Pragmatic trials tend to be tailored to answering questions about specific populations
- But, often the comparator treatment or intervention option is not very specific
  - What is ‘usual care’?
  - How does it vary between trial sites?



Rule 2 – Choose modifiable exposures







Choose exposures that are modifiable (at least theoretically)

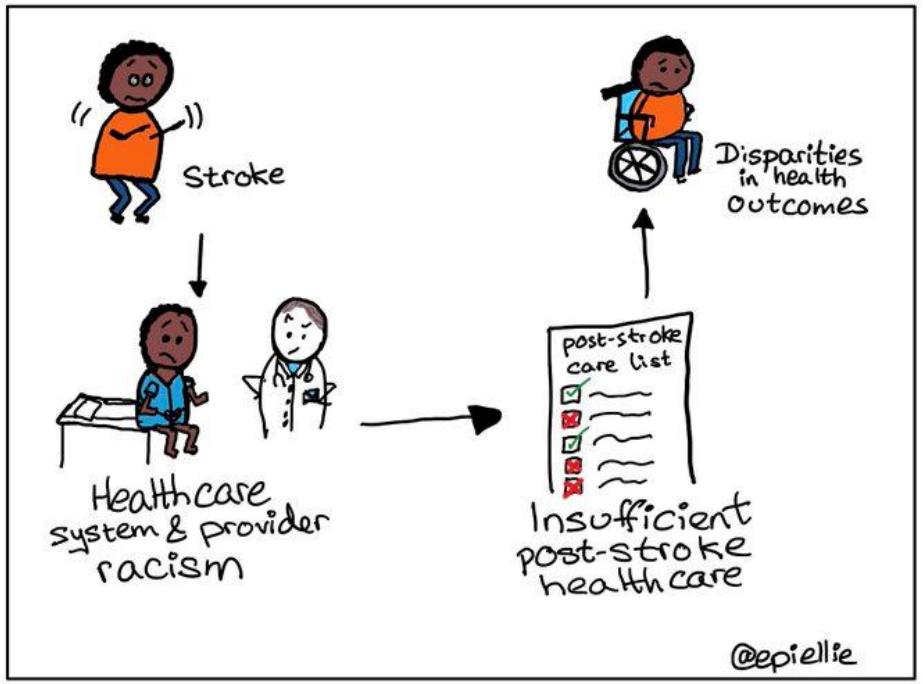
If we don't know how to directly modify something, we probably can't estimate it's causal effect

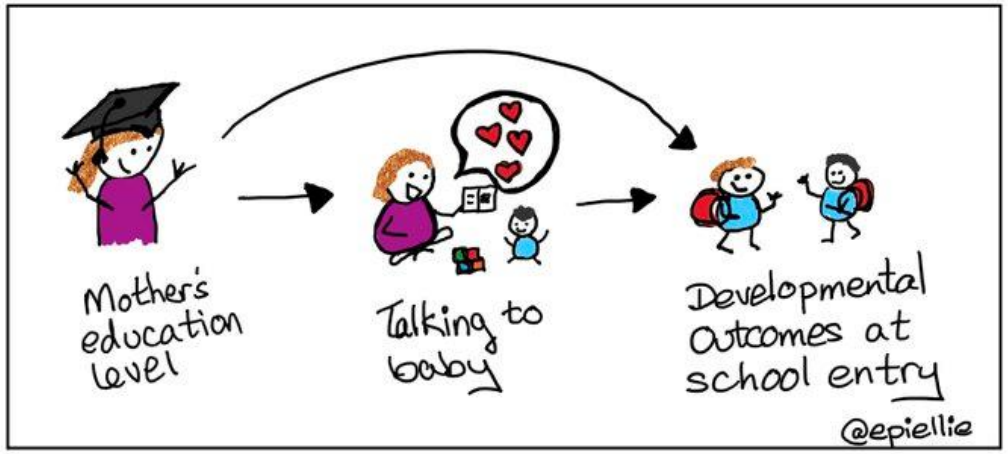
We can only estimate causal effects of well-defined interventions on outcomes we believe can be changed by the intervention

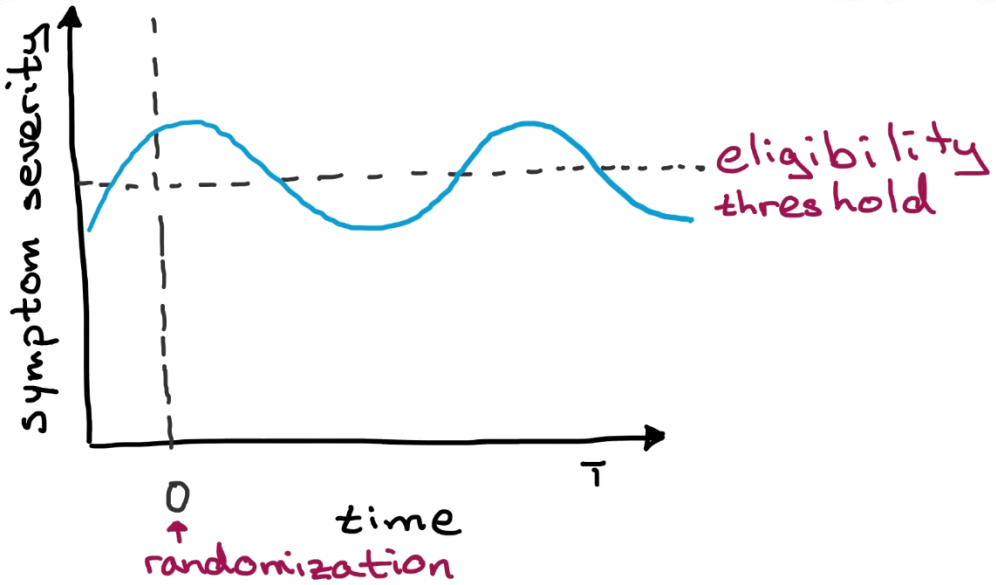
Pragmatic trials and modifiable exposures

- In general, all intervention studies should have a modifiable exposure by default
  - The intervention is assigned & this by definition makes it modifiable
- But, in pragmatic trials, we may have more non-adherence to assignment
  - Non-adherence is not necessarily something we are modifying or can modify.
- Need to consider:
  - What is the level of adherence in each trial arm?
  - What trial activities are directly impacting adherence?

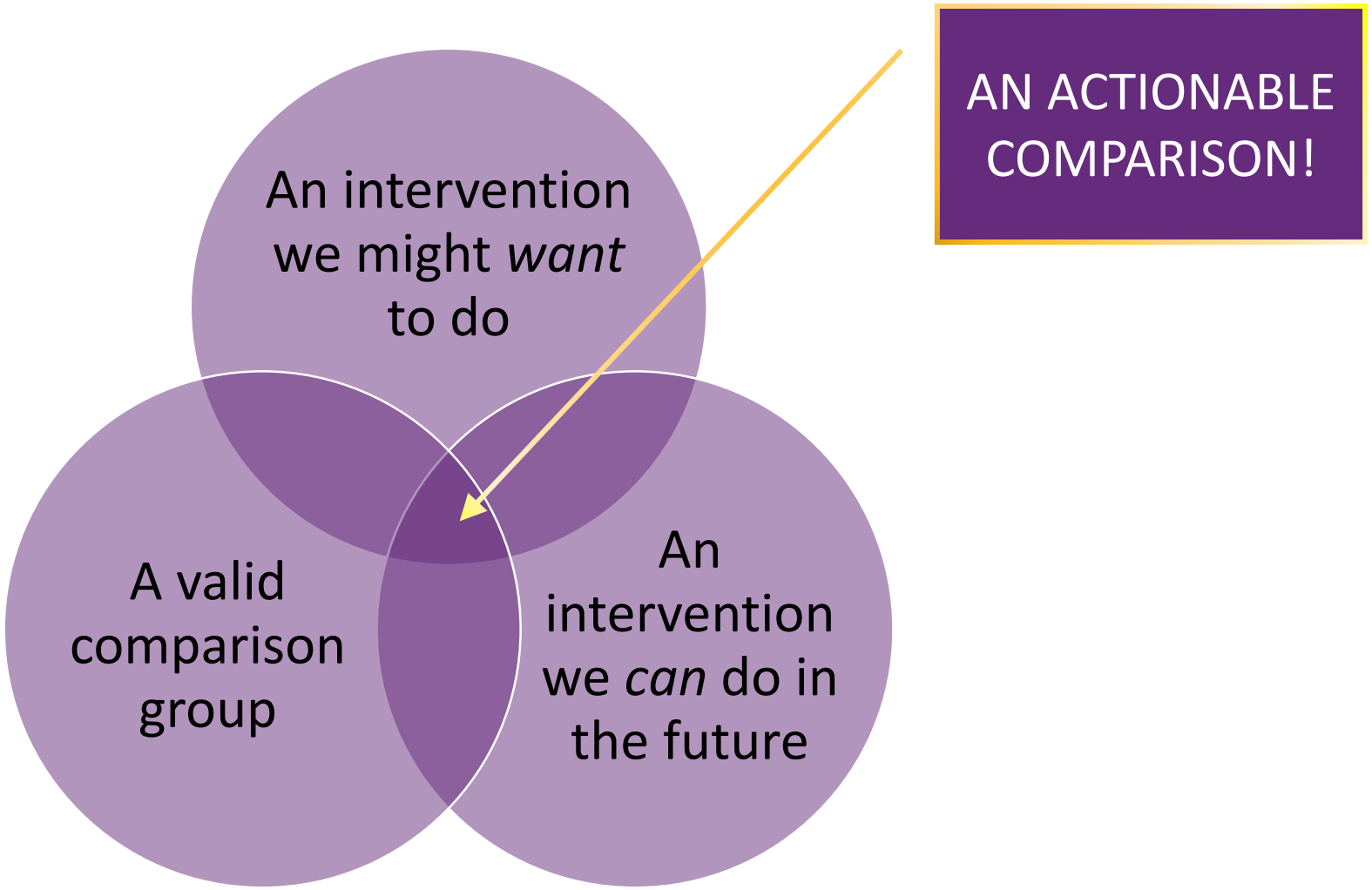
Rule 3 – make your comparisons actionable





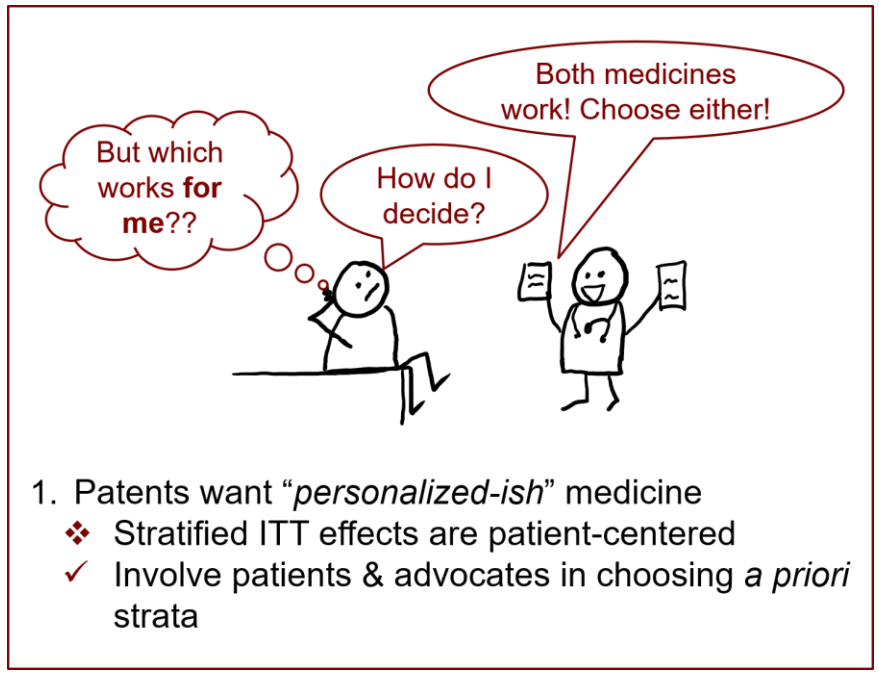


What do we need to know to act on information?



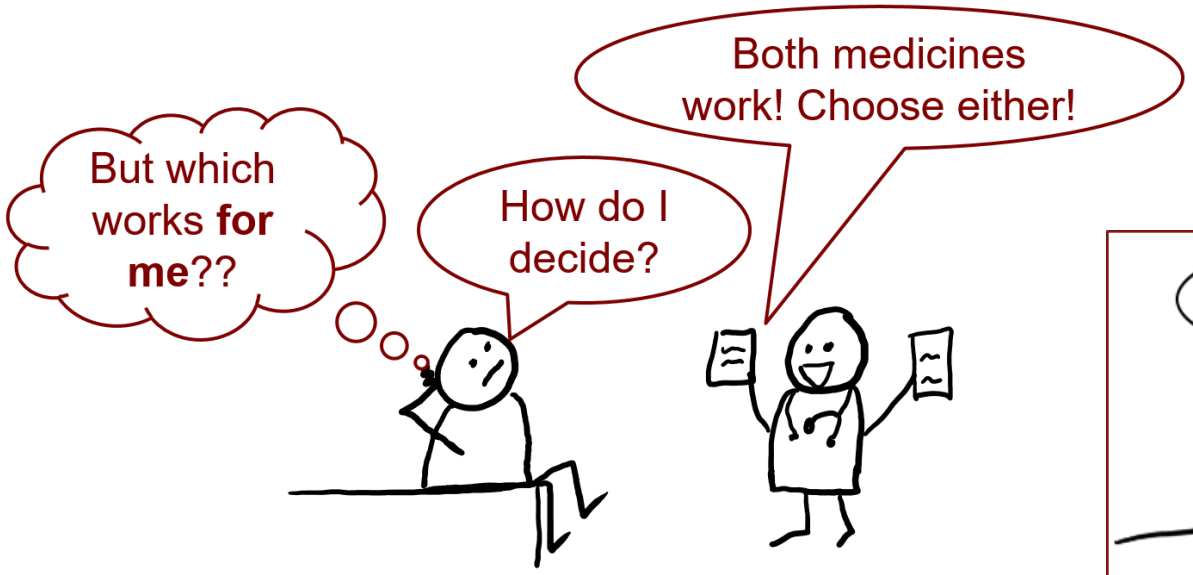
Pragmatic trials are actionable

- Actionable comparisons is the typical goal of pragmatic trials, and a strength that sets them apart from explanatory RCTs
- Actionability can be further improved by involving patients & advocates in choosing categories for stratified effects
- Patients want personalized-ish medicine
  - How will it work for someone like me?

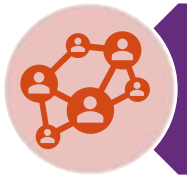


Murray et al. 2018 J Clin Trials, 103:10-21.

Rule 4 – use a relevant comparison, measure, and study design



Relevant questions target decisions we are uncertain about



Does this drug work for the type of patients I treat?

Which of these two drugs work better for patients who've failed first line treatment?

If patients expect to be able to take the drug regularly, will it work better than another option? What if they don't?

Even if two drugs are equally effective, they may have different side effects – how much safer is the new option?

Relevant questions: Non-inferiority comparisons are often 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 et al. 2018 J Clin Trials, 103:10-21.

Choosing a relevant comparison:

Intention-to-treat effect gives an unbiased estimate of effect of randomization…

But,

- Lower bound on the effect of treatment if compared to placebo if non-adherence
- Potentially a problem if assessing outcomes such as adverse events or safety
- When comparing active treatments, ITT vary towards or away from the null if non-adherence

Patients who expect to adhere want to know per-protocol effects

"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 et al. 2018 J Clin Trials, 103:10-21.

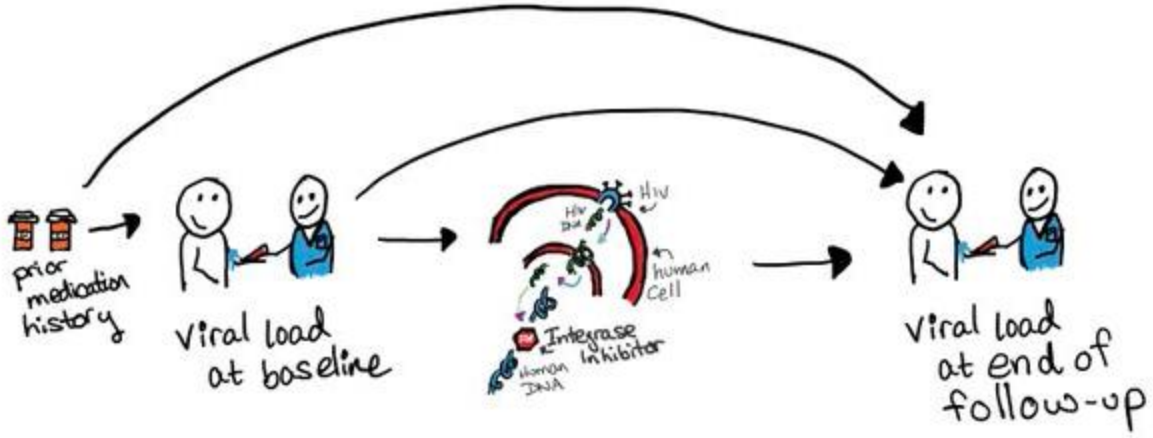
What causal effects can we estimate?

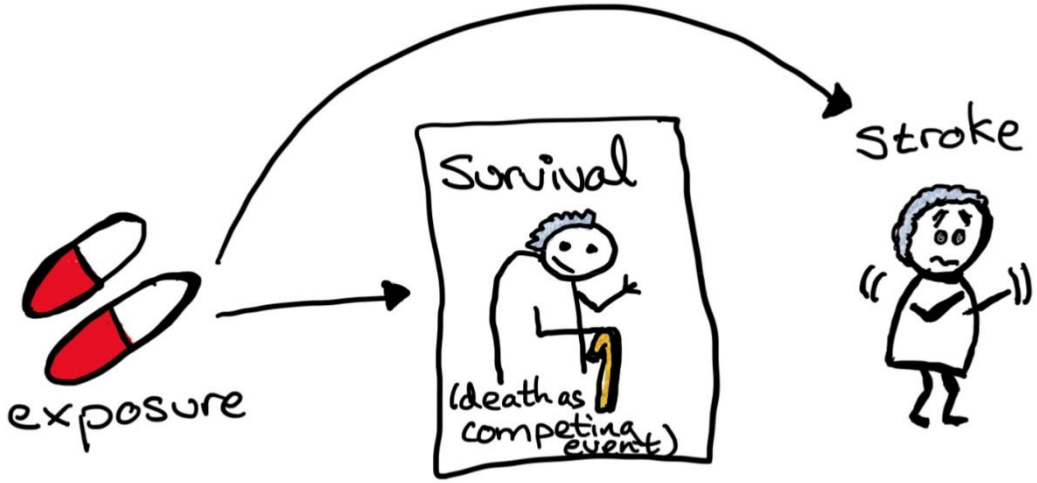
1. Intention-to-treat effects
   - Effect of assignment to treatment ← classical ITT
   - Effect of assignment to treatment had there been no loss to follow-up
2. Per-protocol effects
   - Effect of initiation of assigned treatment
   - Effect of adhering to assigned treatment protocol ← per-protocol & as-treated
   - Effect of receiving point intervention, among the compliers ← randomization as instrumental variable

both sometimes called modified ITT

(Compliers ≠ adherers)

Rule 5 – timing matters





Causation and time are inseparable

Causes must happen before effects	➢ Reverse causation bias
To look at sub-groups, we must use data from before the cause happens	➢ Conditioning on a mediator & collider bias
Other things can happen between causes and effects	➢ Survivor bias & competing events
Intervention is more likely the longer a person is sick or alive	➢ Immortal time bias

Pragmatic trials and timeliness

- Loss to follow-up, non-adherence, and competing events are the biggest time-dependent features of concern for randomized trials, including pragmatic trials
- Loss to follow-up can cause bias even when estimating the Intention-to-Treat effect
  - Avoid bias by appropriately measuring and controlling for loss to follow-up
  - Plan to record characteristics that might be related to loss to follow-up so that you can assess whether you have informative or non-informative censoring

Per-protocol analyses have a bad reputation

[redacted] Following

If you think 'per protocol' analyses tell you the effect of a treatment if it is adhered to, you probably shouldn't be responsible for analysing or interpreting RCTs.

3:01 AM - 27 Mar 2018

6 Likes

1 6

But per-protocol analyses aren't necessarily the same thing as per-protocol effects

Per-protocol analysis answers the question :

"how did trial outcomes differ between those who did adhere to, or recieved, assignment A and those who did adhere to, or receive, assignment B?"

Distinguishing between effects and analyses

Per-protocol effects answer the question:

> "how would trial outcomes differ if everyone adhered to assignment A versus if everyone adhered to assignment B"

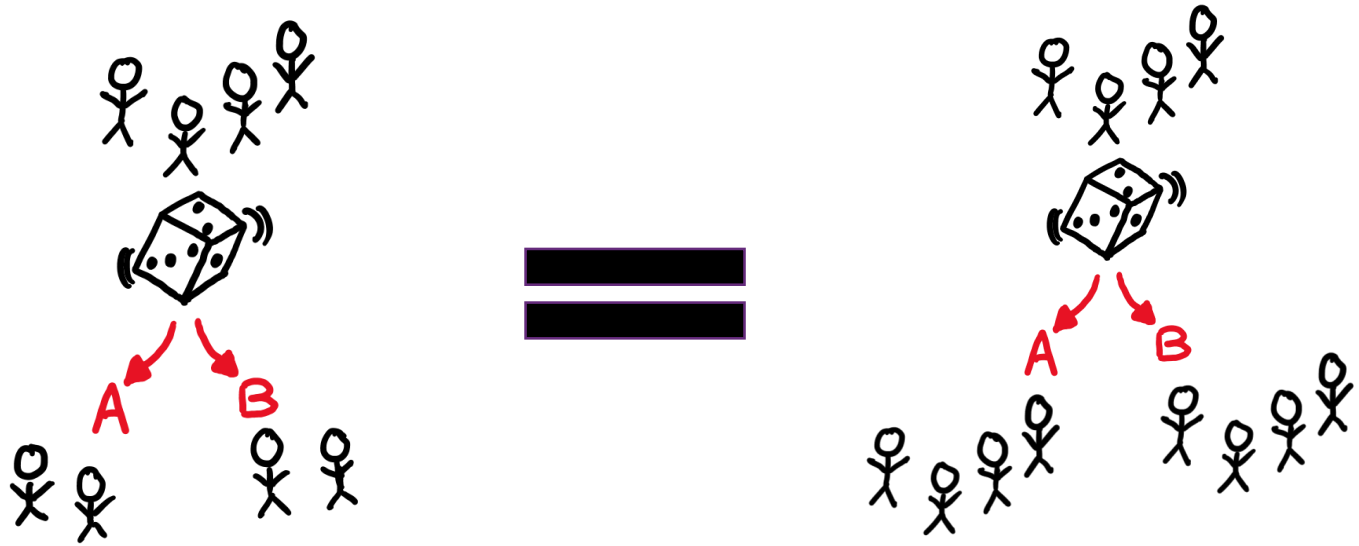
Many ways to answer this question

Appropriate choice may depend on trial design, adherence definition, and confounding

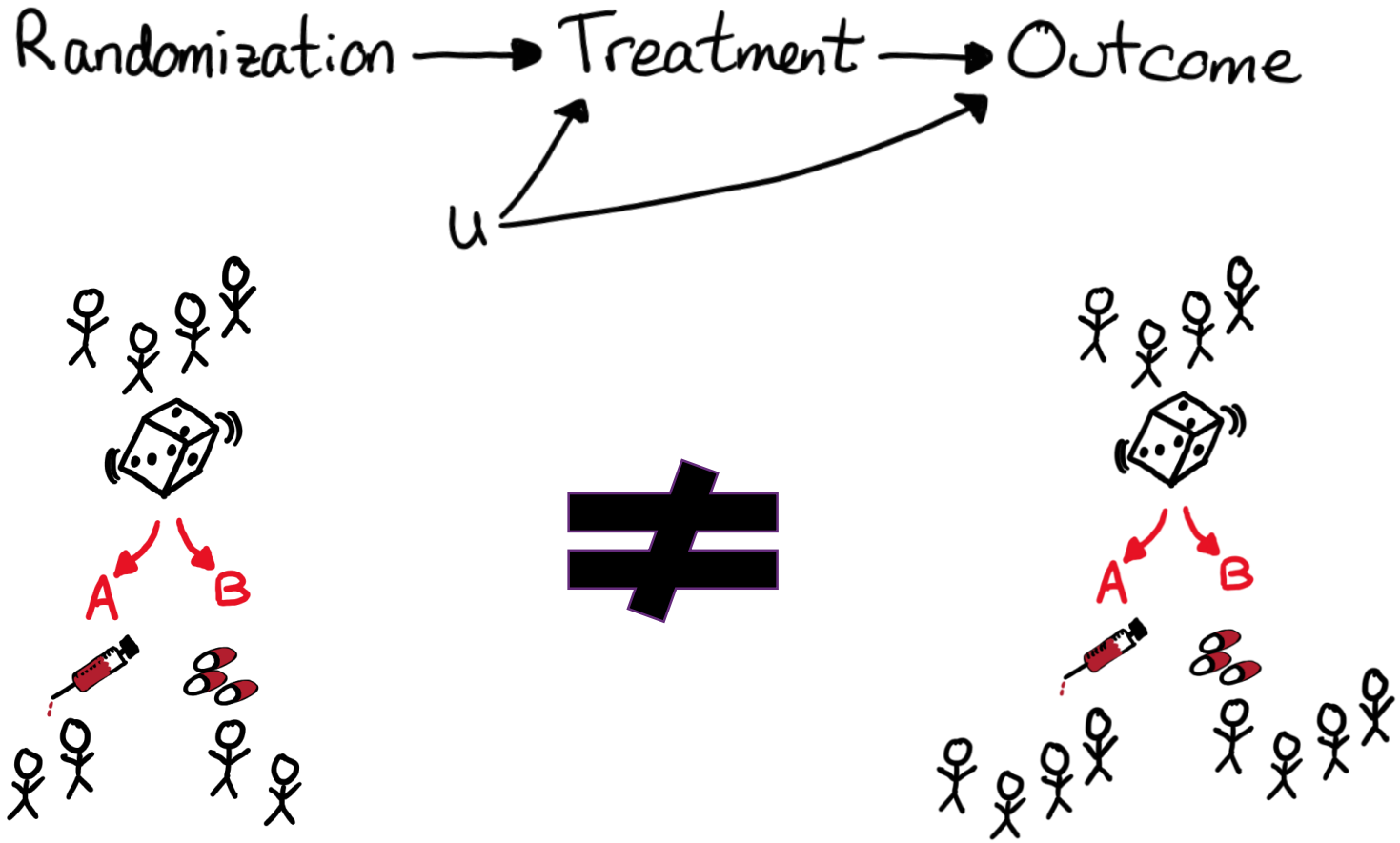
Potential per-protocol analyses

Approach	Description

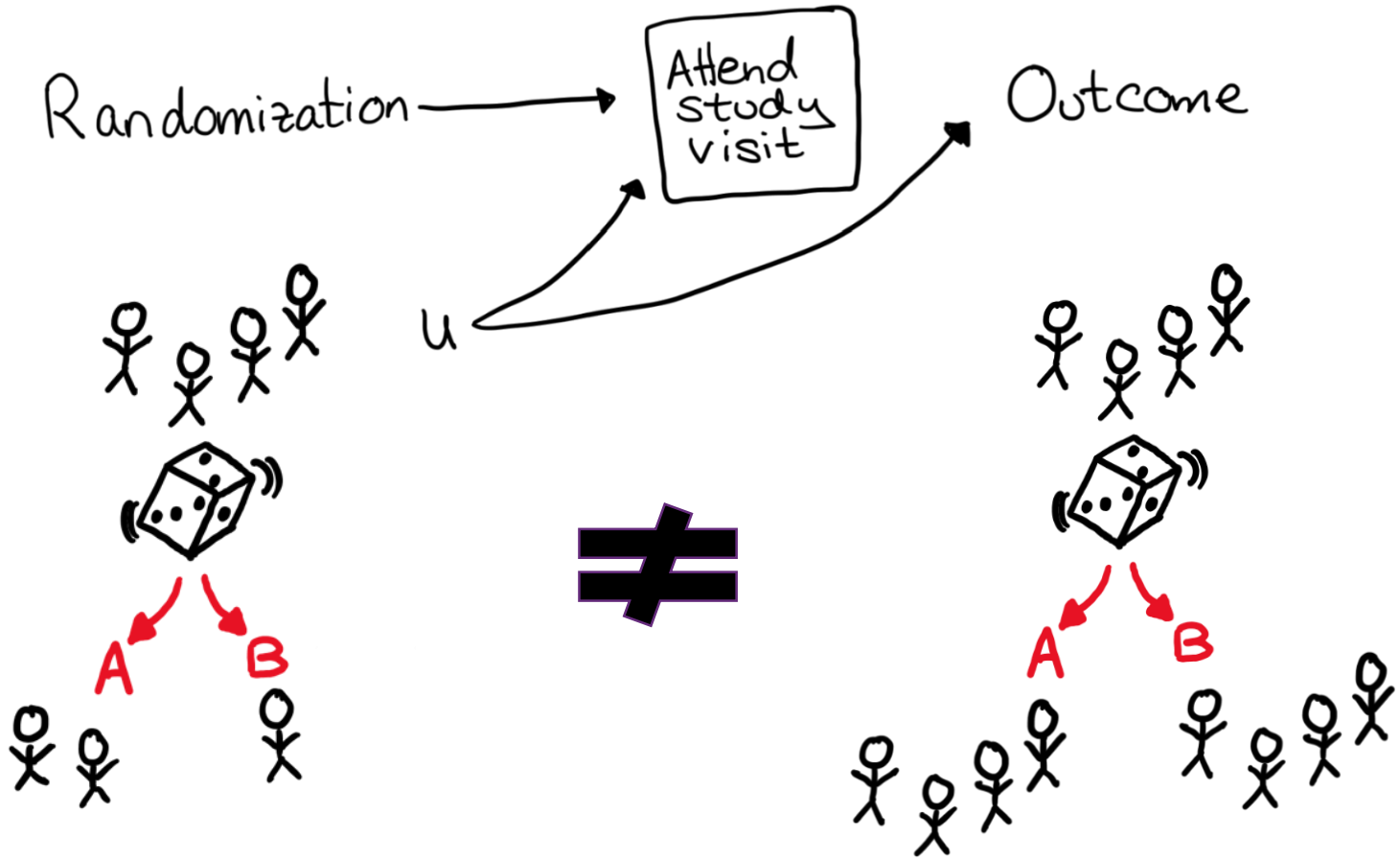
Intention-to-treat effect has no unmeasured confounding for randomization



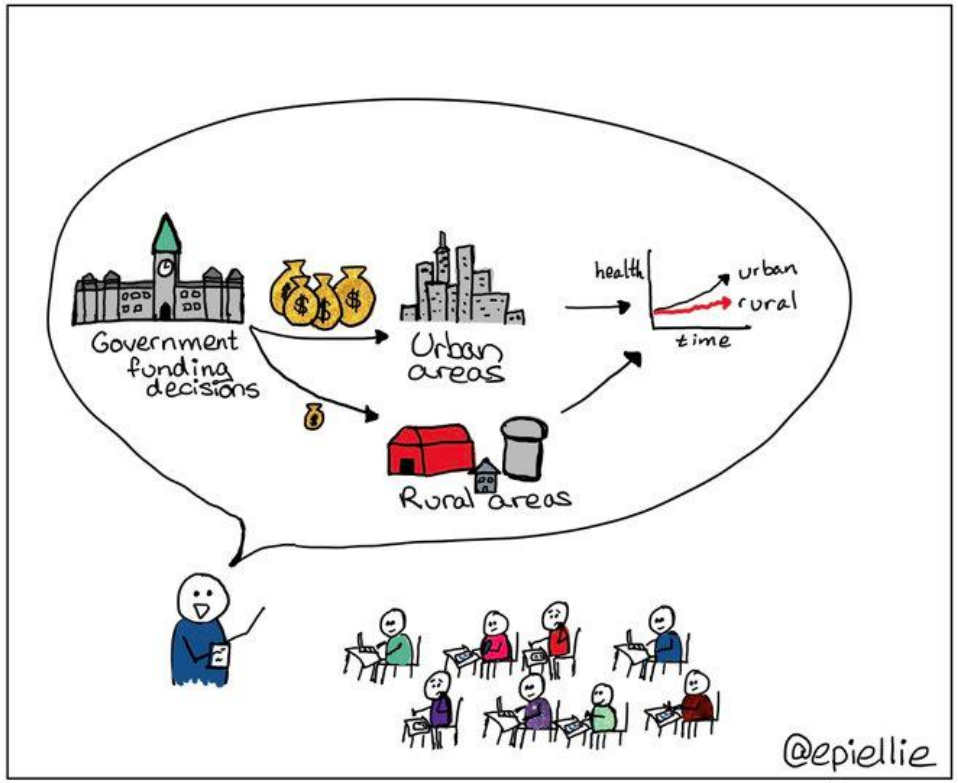
Per-protocol effect almost always requires adjustment for confounding

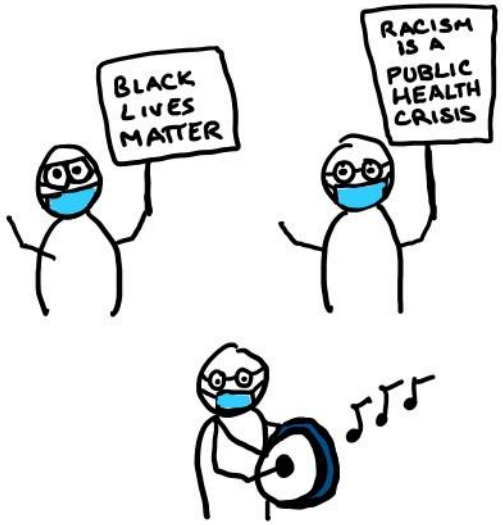


Both ITT & PPE might have bias from loss to follow-up



Rule 6 – design your research to actively promote equity

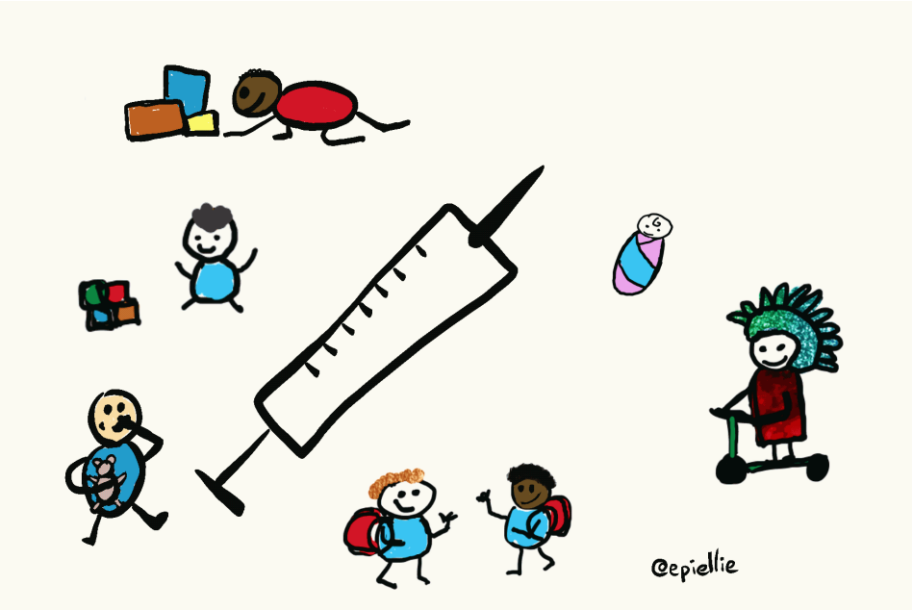




Public health is health for everyone

- Medicine focuses on how best to treat an individual
- Public health is about how best to improve society for everyone
  - Addressing inequities, discrimination, and other barriers to good health is a key component of public health
- When choosing research questions, we should consider how our questions help reduce or promote inequity

Rule 7 – representation matters



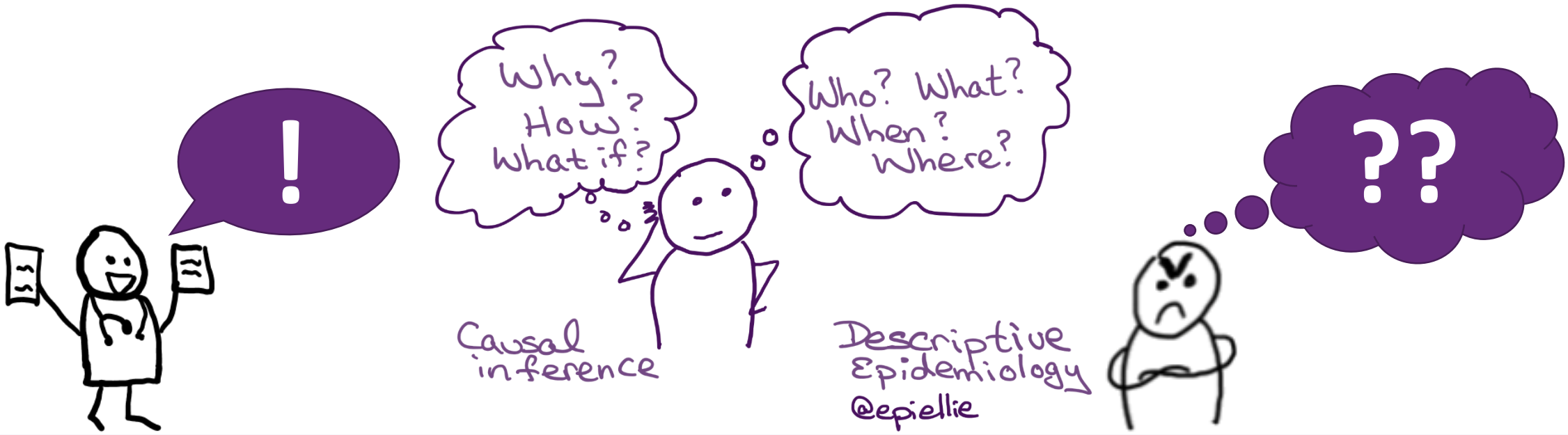
Choose the study population for decision-making relevance

- Representation doesn't necessarily mean all types of people need to be in every study
  - Could mean conducting a study in a specific understudied population or could mean conducting a large study in many types of people
- This is different from *statistical* representation
  - We don't need to have a perfect sample of the underlying statistical population
  - We do need to have scientific information to support decisions for everyone

We got a question, now what?

A two-step causal algorithm:

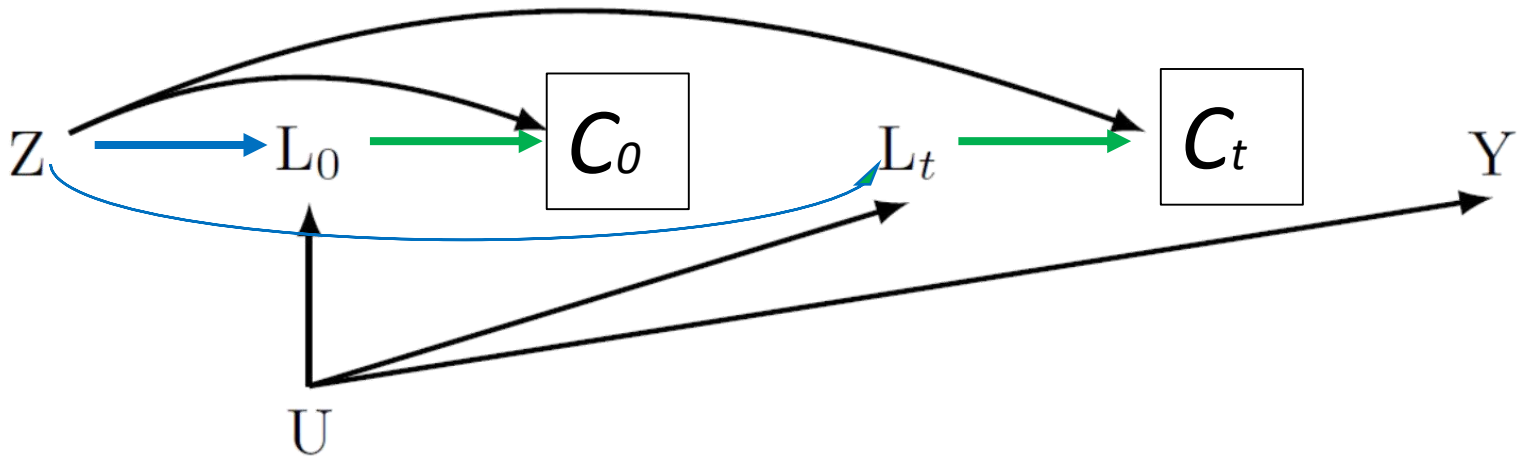
1. Ask good questions
2. Answer them with appropriate methods



First, identify your problem

- Is there loss to follow-up?
  - Does it differ between trial arms?
  - What characteristics of individuals seem to be related to drop-out?
- Is there non-adherence?
  - How much? How does it differ between trial arms? Who adheres and who doesn't?
- Are there competing events?
  - How might these affect interpretation of your answers?

What do we need to adjust for in our ITT estimates?

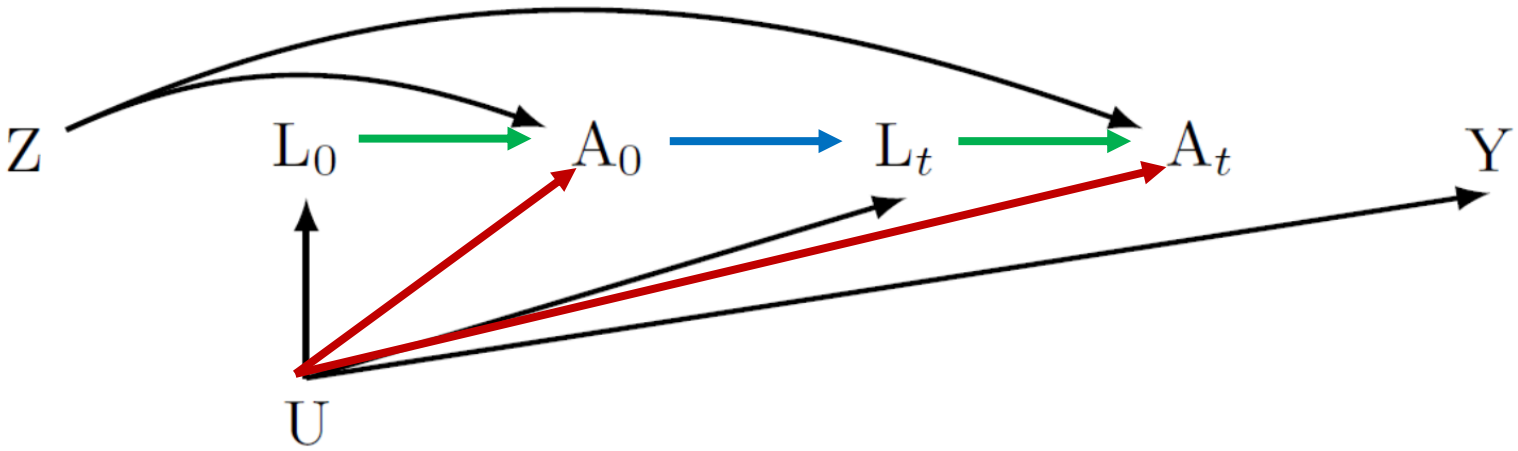


A. Random: No adjustment needed

B. Loss to follow-up confounding by measured covariates

- Adjustment using g-methods may be required if these covariates vary between trial arms
- Otherwise, standard adjustment methods can work

What do we need to adjust for in our per-protocol effect estimates?



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

A quick handshake intro to g-methods

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 are roughly similar to …

Use when you have treatment-confounder feedback….

…and you would normally use these.

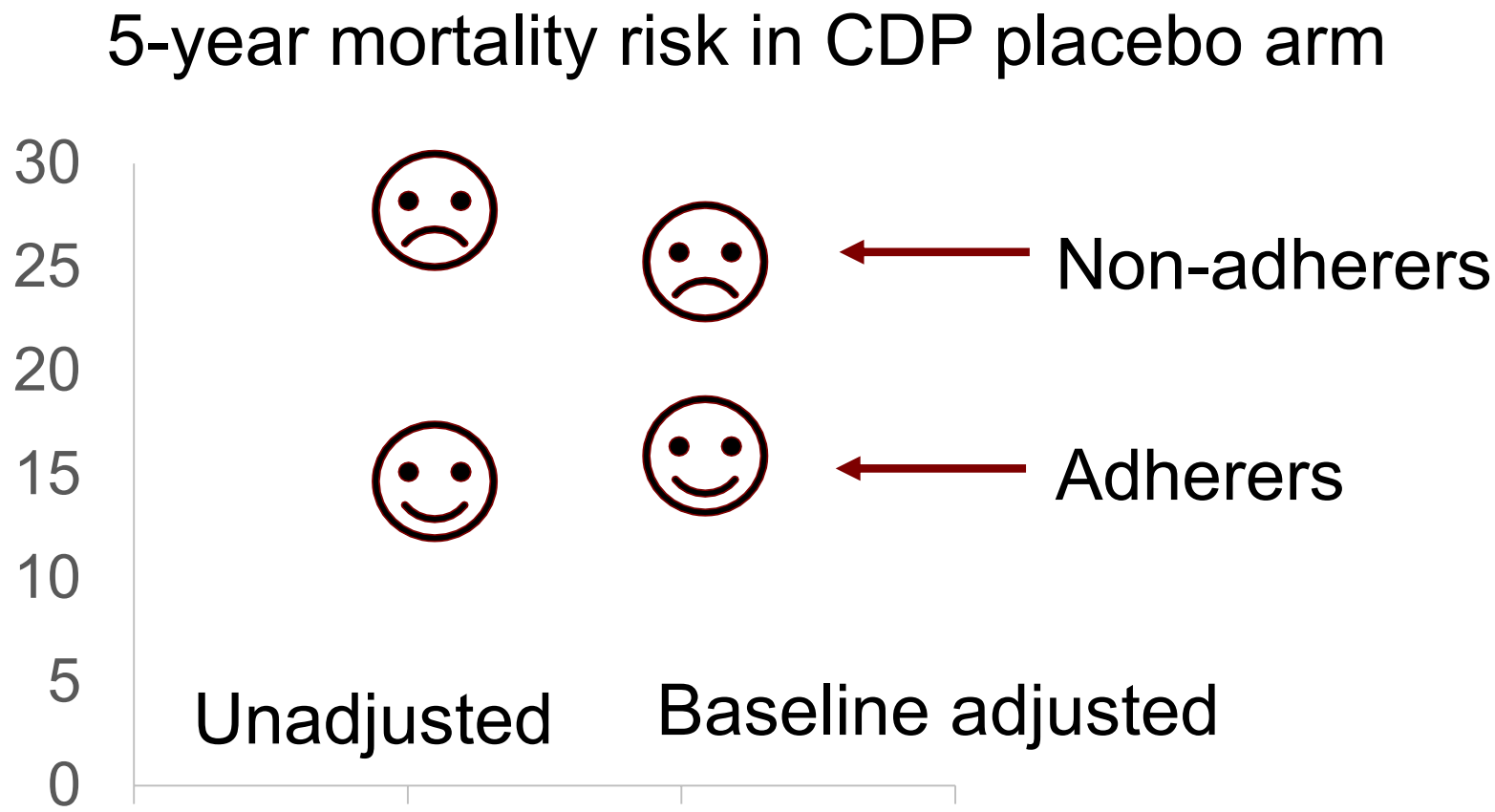
Inverse probability weighting ≈ Propensity scores

(Parametric) G-formula ≈ Standardization

G-estimation ≈ Instrumental variables

Doubly-robust estimation can be used when you would use either IPW or the G-formula

Sounds nice but isn't adherence *intractably* confounded?



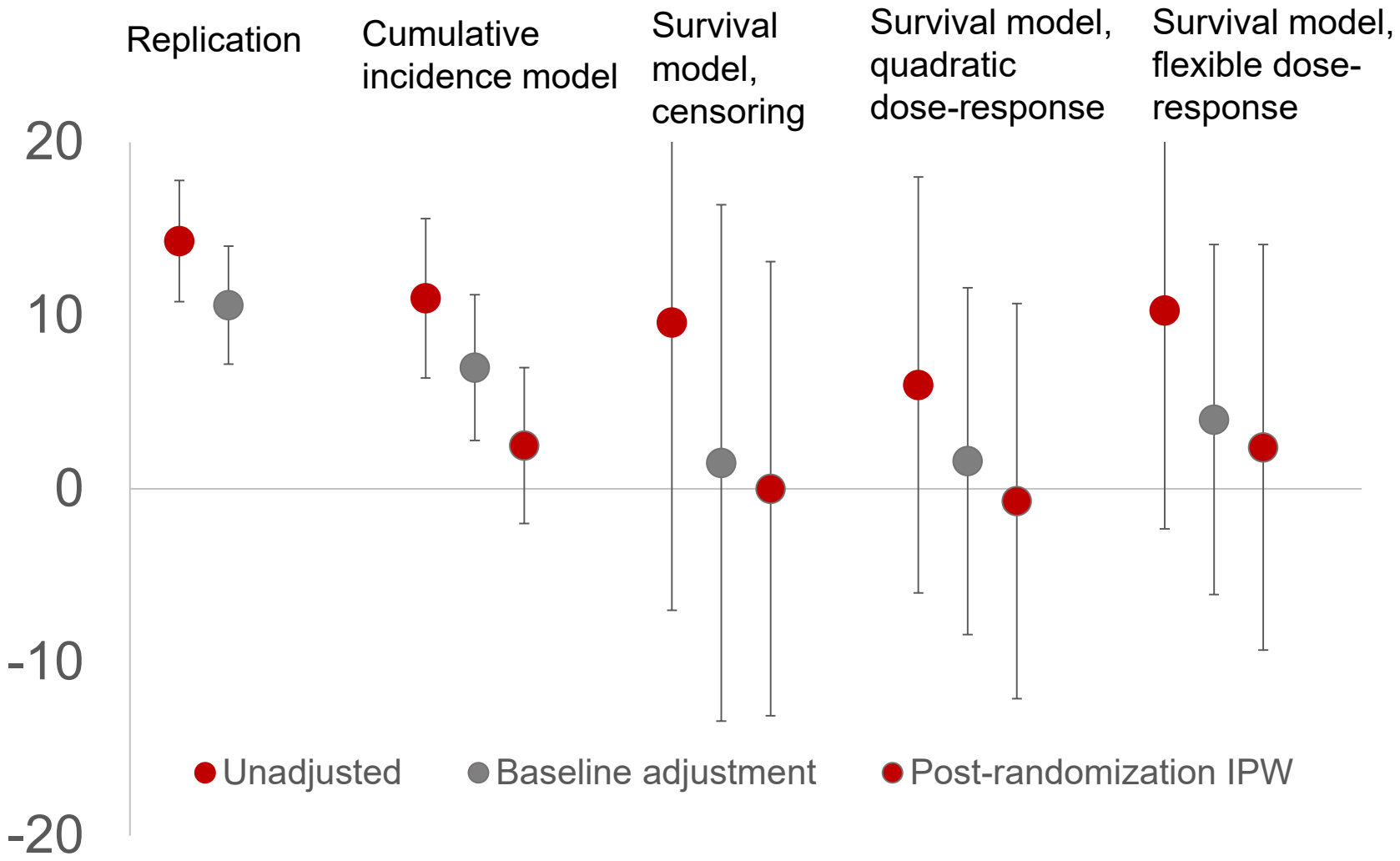
Coronary Drug Project. 1980, NEJM; 303: 1038-41.

Adherence is confounded…

- … but not necessarily *intractably* so!

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

Revisiting the Coronary Drug Project



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

Summary

- Good causal research questions should be SMARTER: specific, modifiable, actionable, relevant, timely, equitable, and representative
- Pragmatic randomized trials are susceptible to bias due to loss to follow-up and non-adherence
  - These biases can be mitigated by appropriate adjustment for confounders.
- When loss to follow-up or non-adherence is random or only affected by measured covariates, standard statistical methods can work
  - If loss to follow-up or non-adherence is affected by prior treatment or adherence, then more complex methods are needed
  - If they are also affected by unmeasured confounders, then stop and call an expert!

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

Questions?

Contact me:

@EpiEllie

ejmurray@bu.edu

https://github.com/eleanormurray

@IMPACTcollab1

IMPACTcollaboratory.org