Evaluation approaches and study designs

GSTTP research mini-course

Corrina Moucheraud, UCLA FSPH

30 March 2021

Learning objectives

- Understand the premise of causal inference and its role in public health and medical research.

- Describe types of study designs, and the strengths and limitations of each for answering causal questions / conducting evaluations.

Discussion

- Discussion about past experiences with evaluations:
  - Have you ever conducted an evaluation?
  - How was it designed?
  - What did you measure?
  - Did you face any challenges?

Types of evaluation designs

- Formative evaluation: will this program be feasible, acceptable, appropriate *[conducted before you implement]*
- Process evaluation: is this program being/has this program been implemented as planned *[conducted during, or sometimes after, you implement]*
- **Outcome evaluation**: did this program change outcomes in the target population *[conducted after you implement]*
- **Impact evaluation**: did this program achieve effectiveness towards ultimate objectives *[conducted after you implement]*

Did SMS reminders increase uptake of HPV vaccine?

= outcome

Did the burden of cervical cancer decline?

= impact

Causation – Why do we care?

- In much of public health, clinical, translational science research, we want to be able to say "action X caused outcome Y"
  - When people took azithromycin, their sinus infection cleared faster *(… than it would have without the azithromycin)*
  - When people received our text messages 3 times per week, they were more likely to quit smoking *(… than if they hadn't received those messages)*
  - When people participated in our training course, their knowledge about gender-based violence increased *(… more than it would have without the course)*

- *Italics = the counterfactual*

Counterfactuals are unicorns

- The counterfactual <u>does not occur</u> – it is a hypothetical state of the world
  - Either people got the azithromycin, or they didn't. There is <u>no one who both did and did not</u>.

- Our goal is to therefore estimate the counterfactual
  - Find a group that emulates the counterfactual state of the world
    - Exactly like the people who got the azithromycin, in every way except the azithromycin receipt

Briefly: why is causation so important?

- There are lots of things going on in the world all the time! – some of these may influence your outcomes
  - Some sinus infections clear up without antibiotics
  - There are various PSAs & other smoking cessation interventions & policies
  - Awareness of gender-based violence is going up for everyone ("secular trend")

- Usually we've spent a lot of time, money, effort on an intervention – so we want to be sure that we can demonstrate an effect of that intervention without this "noise"

{CM opinion: we've all become a little too obsessed with causation…}

Let’s revisit those earlier questions

- When people took azithromycin, their sinus infection cleared faster ~~*(… than it would have without the azithromycin)*~~ *(… than very similar people who did not receive azithromycin)*
- When people received our text messages 3 times per week, they were more likely to quit smoking ~~*(… than if they hadn’t received those messages)*~~ *(than very similar people who did not receive the texts)*
- When people participated in our training course, their knowledge about gender-based violence increased ~~*(… more than it would have without the course)*~~ *(than the knowledge of people who did not participate in the training)*

{CM aside: there may also be unintended consequences, or spillovers… not addressing that today but also merits consideration}

Ok but if causation is important, but counterfactuals are hypothetical – what now??

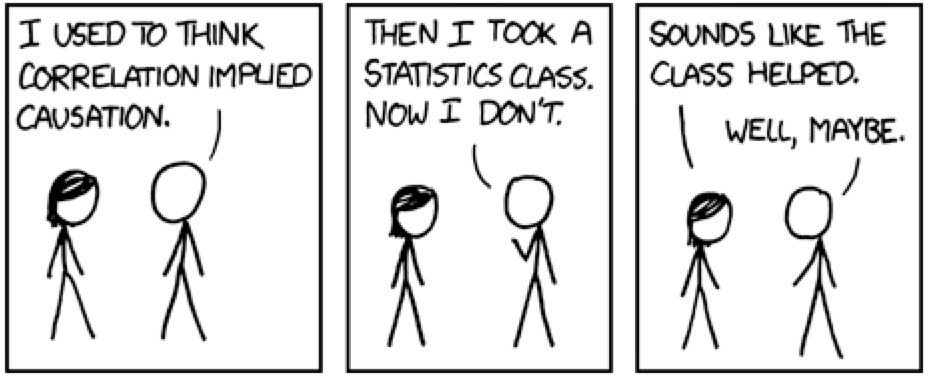
- Causal inference is an analytic approach, with associated statistical tools, for estimating the causal effect *{CM: emphasis on "inference"!}*
- Gold standard: randomized clinical trials
  - Because treatment is assigned at random, we have every reason to think that people who are exposed to the intervention are identical in all other ways to those who are not exposed ("control" group)
- But what about things that are impossible or unethical to randomize?
  - Other options ("quasi-experimental"): natural experiments, regression discontinuity, instrumental variables…. To compare intervention & "comparison" groups
    - Happy to talk more about these but not today's focus

So usually we do non-experimental evaluations

Most common types:

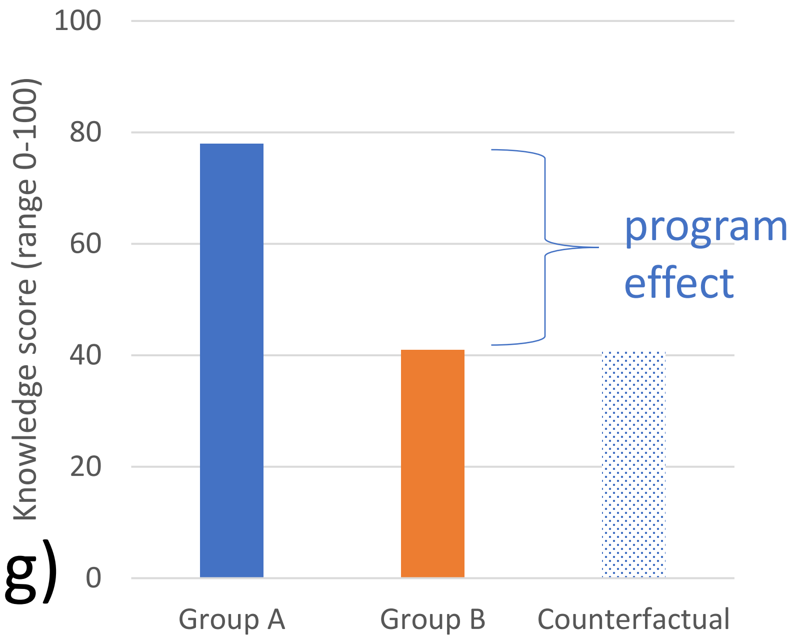
- Comparison
- Pre-post
- Difference-in-difference

- Always trying to avoid “mere” correlation because we want to be able to attribute any effects to our intervention



Credit: xkcd

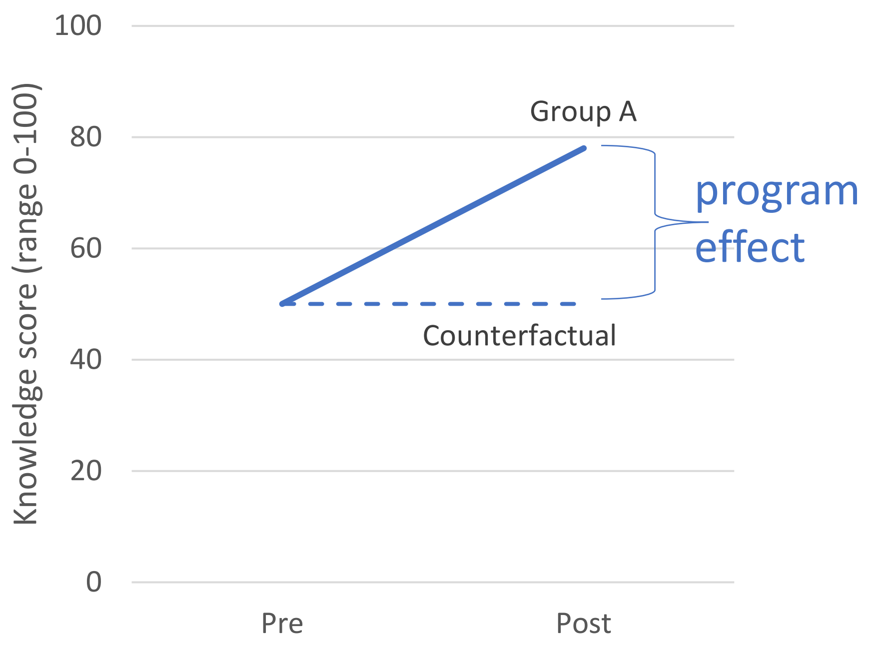
Comparison



- Look at the knowledge score in group A (did the training) and compare it to the knowledge score in group B (did not do the training)

- If groups A & B are very *(very very)* similar, this can be informative
  - You can analyze the data to "control for" differences – for example, more people in group A went to college; or, people in group B are younger
  - But, what if there are things about the groups that you haven't observed?
  - And, what if people could opt in or out of the program & this is associated with the intervention – like if people who felt the training was too hard opted out

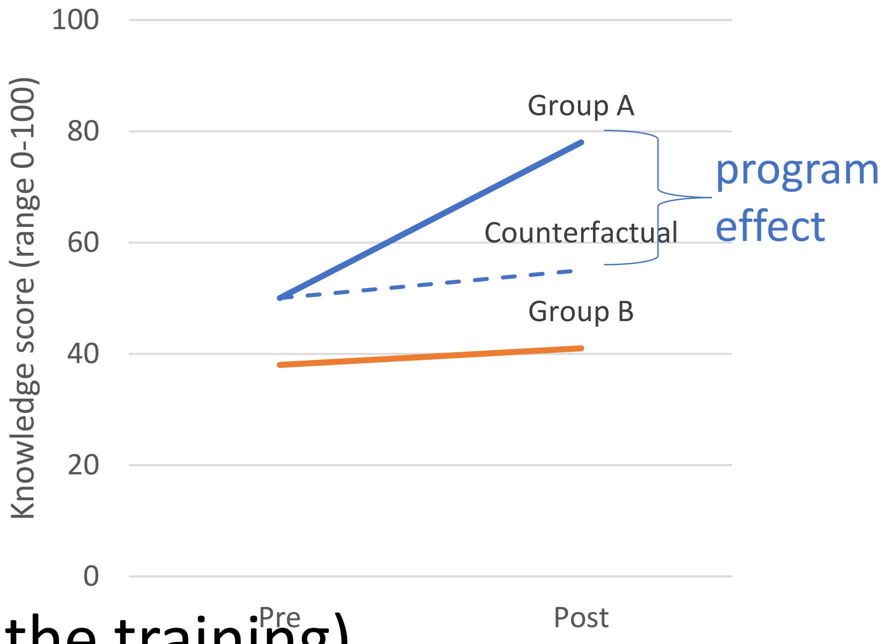
Pre-post design



- Look at the knowledge score in group A before and after they complete a training

- Now you don't have to worry about differences between groups (a person's pre score is the counterfactual – assumed to be what their knowledge would have been in the absence of the training) but what if their knowledge had already been on the up-swing ("pre-trend")?
  - And, what if group A is not representative of other people?
  - Also, same challenge about opting out

Difference-in-difference



- Look at how the knowledge score in group A (did the training) changed from pre to post, and compare this it to the change in knowledge score over this same period in group B (did not do the training)

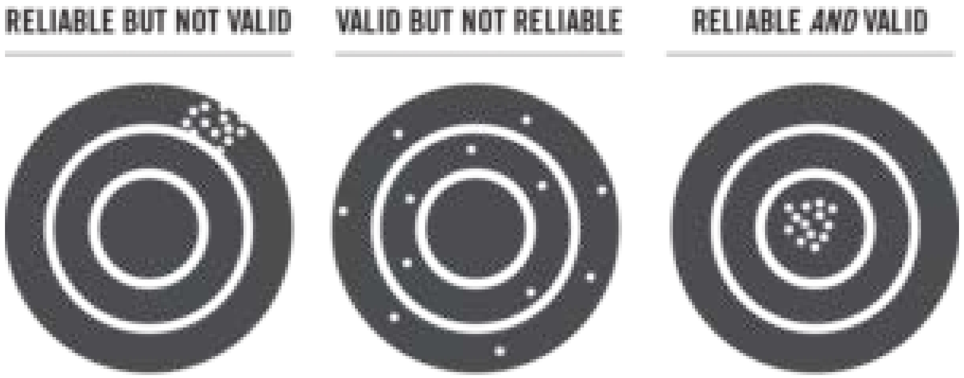
- Now, we are "controlling for" any secular changes, as well as any differences between the groups that might affect their change over time besides the intervention exposure – i.e., we are interpreting the slope of the Group B line as the counterfactual
  - Still does not deal with pre-trends
  - Still does not deal with opt-in/-out
  - And does not address anything that might affect the groups during this period unassociated with the intervention

Key take-aways:

- We aspire to tell a causal story when we do an impact (or outcome) evaluation
  - Because we implemented this intervention, things changed
- Key to a causal story is a counterfactual
  - What would have happened to <u>these exact same people</u> if they had not been exposed to the intervention
- Counterfactuals are impossible so we try to emulate them
  - Gold standard: experimental designs like drug trials → but this is not always feasible, ethical, possible
  - Next best: “quasi-experimental” approaches
  - Next-next best: observational studies → all have pros & cons, so need to be cognizant of these when you design & interpret (& consume) these

- NOTE: evaluation design should be specified up-front (before implementation)!

A final note on validity & bias…



- In statistics classes, you learn concepts of validity & reliability
- We will talk later about reliability (are we measuring things with precision & meaning); <u>evaluation design is primarily concerned with validity</u>
  - Internal validity: are we observing a “true” effect (or is it an artifact of something we’ve done)
  - External validity: is the effect we observe generalizable to anybody else
- When we talk about “bias,” this primarily affects internal validity
  - Selection bias
  - Non-response bias
  - Survivorship bias
  - Etc.

Discussion

- Are you doing an evaluation for your STTP project? What type of evaluation (formative, process, outcome, impact)? What design are you using? What are some of the strengths & weaknesses you can anticipate?

More resources

- BetterEvaluation collaborative: https://www.betterevaluation.org/en

- Numerous books including:
  - Causal Inference: The Mixtape (https://mixtape.scunning.com/)
  - Mostly Harmless Econometrics (https://www.mostlyharmlesseconometrics.com/)
  - Causal Inference: What If (https://www.hsph.harvard.edu/miguel-hernan)
  - Impact Evaluation in Practice (https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice)
  - And many more!