Translating causality into practice:

Causal questions and analysis in applied epidemiology

Jonathan Snowden, PhD
Kelly Reavis, MPH
OHSU-PSU School of Public Health

17 June, 2019
SPER
Advanced methods workshop

Objectives

1) Discuss advanced epidemiologic methods, putting the question first (i.e., before data and analytical considerations).
2) Engage deeply with the process of *formulating causal questions*.
3) Based on *features of the question*, provide an overview of some advanced methods and when they are indicated/useful.
4) Introduce one high-priority question from maternal health research:
    a. Generally: Causes of severe maternal morbidity (SMM), such as maternal BMI, obstetric procedure use (e.g., epidural, cesarean), labor duration
    b. Specifically: What are the mechanisms explaining the effects of BMI on postpartum hemorrhage (PPH)?

Objectives (2)

5) Encourage you to use causal diagrams to explain the inter-relationships between the content of this question (i.e., “causal structure”).

6) Motivated by our question, to explain the basis of causal mediation analysis.

7) Demonstrate the mechanics of a method (IPTW) and explain its rationale for mediation.

   a. Provide an opportunity for you to implement it on simulated data.

8) Connect you with further resources for formulating causal questions and analysis plans.

Workshop agenda

I. Causal questions and assumptions (corresponds to Snowden *JMWH* 2018)

II. Selecting and framing a maternal health question
   - Causal diagram activity
   - Basic data analysis activity

III. Methods and code demonstration (corresponds to Leonard *PPE* 2019)
   - Code demonstration and activity

I: Causal questions and assumptions

Agenda

- Defining CI, avoiding misconceptions
- Notation
- Assumptions
- 3-step process:
  - Step 1: Formulate question
  - Step 2: Assess data
  - Step 3: Design analysis
- Break

Agenda

- Defining CI, avoiding misconceptions
- Notation
- Assumptions
- 3-step process:
  - Step 1: Formulate question
  - Step 2: Assess data
  - Step 3: Design analysis
- Break

Defining causal inference

- Causal inference is a hot topic.
- Yet there's no consensus on how to define CI. Definitions are:
  - Dynamic
  - Disputed
  - Diverse
- **What's clear: causal inference is important.**
  - Most health questions are causal, or will lead to causal questions.

Common CI misconceptions, 1-2

1. CI is a method or set of methods.

- Or, CI requires a given method.
- E.g., DAGs, PS, g-methods, IVs …
- A method does not on its own confer a causal interpretation onto a calculated association.

2. Causality can be inferred in a single study.

- I disagree.
- Causation is inferred across a body of studies (Broadbent 2016)
  - But, we may consider approaches for inferring effects of a more causal nature in a given study.

Common CI misconceptions, 3-5

3. Causal questions are more important than ("better than") non-causal questions.

- There is a taxonomy of research questions; question in each category are important.
- Consider: descriptive, predictive, and causal questions.
- Each is useful at different stages of research.

4. Effective policy actions must be based on sound causal knowledge.

- Consider: Effective action was taken to curb early HIV epidemic, based on mixed/faulty causal basis

5. Sound causal knowledge will invariably lead to improved health.

AMERICAN
Journal of Epidemiology
Formerly AMERICAN JOURNAL OF HYGIENE

VOL. 129 MARCH 1989 NO. 3

Reviews and Commentary

AN AUTOPSY OF EPIDEMIOLOGIC METHODS: THE CASE OF "POPPERS" IN THE EARLY EPIDEMIC OF THE ACQUIRED IMMUNODEFICIENCY SYNDROME (AIDS)

JAN P. VANDENBROUCKE[1] AND VÉRONIQUE P. A. M. PARDOEL[2]

Common CI misconceptions, 6-7

6. Causal inference in new.

- Causal thinking dates back to antiquity.
- Even current epidemiologic methods for CI are not new.

Method	Era
Causal diagrams	1920s (S. Wright)
IVs	1920s (P. Wright)
IPW	1950s (Horvitz-Thompson)
PS	1983 (Rosenbaum)
G-methods	1986 (Robins)

7. Causal thinking is synonymous with counterfactual thinking and models.

- Restricted potential outcomes approach is 1 causal model.
- Others: causal pies; Hill considerations

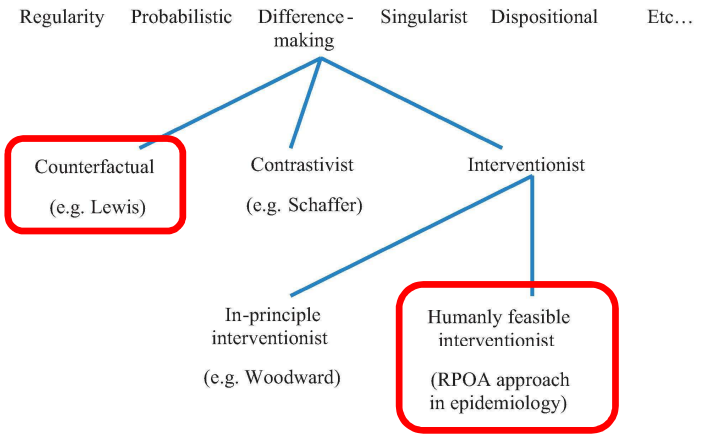


Figure 1. Fitting the restricted potential outcomes approach (RPOA), as advocated in epidemiology, in a family tree of theories on causality.

Defining causal inference

- We've covered some of what CI is not.
- What is it?

In reality, causal inference is a process.

- It is a multi-disciplinary scientific undertaking that unfolds over time.
- It involves triangulating between various findings, sources of evidence, and ways of knowing.

Counterfactual causal inference is our focus.

- It uses specific notation that we will discuss.
- It also rests on 3 assumptions that we will cover.

Agenda

- Defining CI, avoiding misconceptions
- Notation
- Assumptions
- 3-step process:
  - Step 1: Formulate question
  - Step 2: Assess data
  - Step 3: Design analysis
- Break

Terminology and concepts

- **Treatment:** synonymous with exposure
- **Observed data:** Real data that have been collected or will be.
- **Unobserved data:** Hypothetical data (i.e., not real data) that might have been observed, if circumstances were different.
  - These data are *counterfactual*.
  - We can think of them as *missing data.*
- Note that for each individual at any given time, we can only observe one setting for each variable we collect.
- **Counterfactual framework** (synonymous with "potential outcomes framework"):
  - One framework for CI in epidemiology.
  - Assumes the existence of unobserved data that might have been observed if, *contrary to fact*, things had occurred differently.
- **Intervene, set, do:** The process of hypothetically changing an observed value of a study variable to another, often unobserved value.

Notation

- **Notation:**
  - ***Y*** is the random variable for observed outcome
  - ***A*** is the random variable for observed treatment
  - ***W*** is a confounder
  - $\overline{\boldsymbol{W}}$ is a vector of confounders, expands to:
    - W^1: maternal race / ethnicity
    - W^2: maternal age
    - W^3: maternal insurance status
    - etc…

Notation

- **Notation:**
  - ***Y*** is the random variable for observed outcome
  - ***A*** is the random variable for observed treatment
  - ***W*** is a confounder
  - $\overline{\boldsymbol{W}}$ is a vector of confounders, expands to:
    - W^1: maternal race / ethnicity
    - W^2: maternal age
    - W^3: maternal insurance status
    - etc…
  - Capital letters are random variables:
    - ***Y***: postpartum hemorrhage
  - Lowercase letters are specific realizations / values of these variables:
    - ***y*** = 0 (no PPH), or
    - ***y*** = 1 (PPH occurred)

Notation for observed data

- **Notation:**
  - Pr(A=1) Probability of our (binary) exposure in our sample.
  - E(Y) Mean of our observed outcome in our sample.
  - E($Y|A$=1) Mean outcome among exposed people.
  - E($Y|A$=0) Mean outcome among unexposed people.
  - E($Y|A=a, \bar{W} = \bar{w}$)

    Mean outcome, conditional on exposure and confounders.

    (This is what we model most frequently in regression models.)

$$E(Y|A=a, \bar{W} = \bar{w}) = \beta_0 + \beta_1 * A + \overline{\beta_2} * \bar{W} + \varepsilon$$

  - E($Y|A$=1) – E($Y|A$=0) Risk difference
  - E($Y|A$=1) / E($Y|A$=0) Risk ratio

Notation for observed data

- **Notation:**
  - Pr(A=1) the probability of our (binary) exposure in our sample.
  - E(Y) the mean of our observed outcome in our sample.
  - E($Y|A$=1) Mean outcome among exposed people.
  - E($Y|A$=0) Mean outcome among unexposed people.
  - E($Y|A=a, \bar{W} = \bar{w}$)

    Mean outcome, conditional on exposure and confounders.

    (This is what we model most frequently in regression models.)

$$E(Y|A=a, \bar{W} = \bar{w}) = \beta_0 + \beta_1 * A + \overline{\beta_2} * \bar{W} + \varepsilon$$

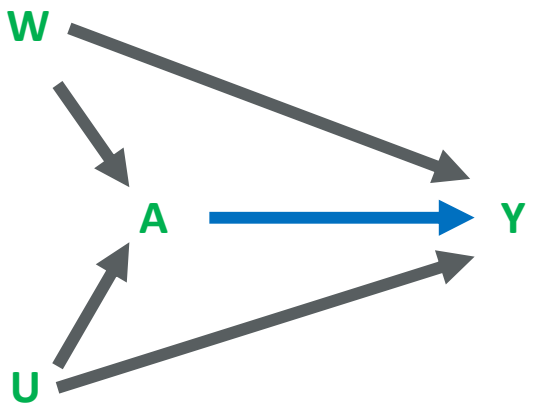
  - E($Y|A$=1) – E($Y|A$=0) Risk difference
  - E($Y|A$=1) / E($Y|A$=0) Risk ratio

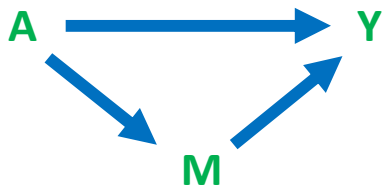
RR = (a/(a+b)) / (c/(c+d))
RD = (a/(a+b)) - (c/(c+d))

	Yes PPH	No PPH
Yes Obesity	a	b
No Obesity	c	d

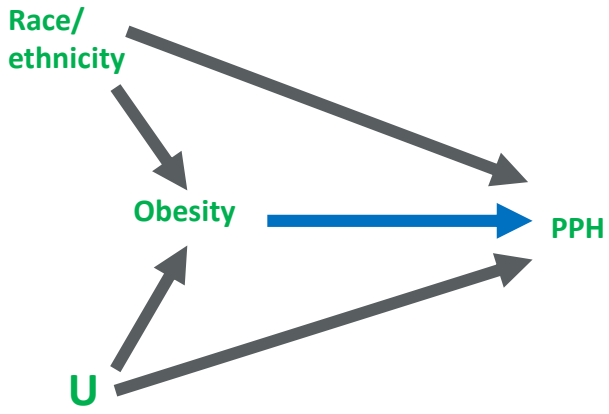
Notation

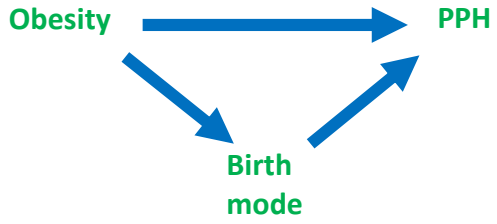
- **<u>Notation (concepts introduced later):</u>**
  - ***U*** represents unknown/unmeasured confounders
    - In observational studies, we assume there will always be ***U*** present
  - ***M*** is a mediator/causal intermediate
  - ***S*** represents selection/retention into the study sample (or attrition from it)

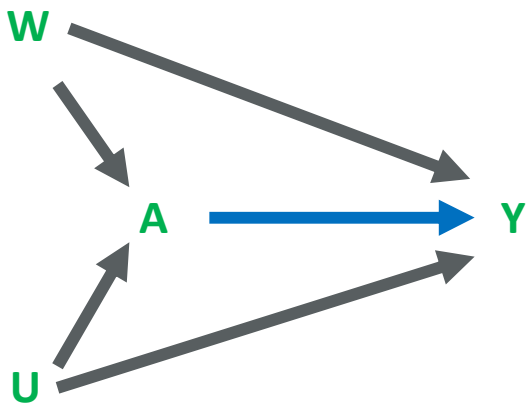


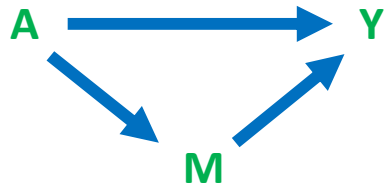


Notation


Race/
ethnicity
Obesity
PPH
U


Obesity
PPH
Birth
mode


W
A
Y
U


A
Y
M

Notation: Counterfactual data

Variable	Setting
A (obesity)	1
M (cesarean)	1
Y (PPH)	Yes

- **Observed and counterfactual data:**
- ***Observed data:*** *A=1; Y=1*
  - Participant has obesity and had a postpartum hemorrhage (PPH)
- ***Counterfactual exposure:*** Set(*a=0*)
  - Set exposure to be non-obese
- ***Counterfactual outcome:*** $Y_{a=0}$
  - This is the unobserved, counterfactual outcome when setting exposure to be non-obese (i.e., *Y,* given Set(*a=0*))
  - The counterfactual outcome is a function of counterfactual exposure:
    - You observe *A* and *Y*
    - You set exposure to be a give value (i.e., Set(*a=0)*)
    - The outcomes take the value of $Y_{a=0}$, based on the counterfactual exposure.

Counterfactual exposures, outcomes

- **Using counterfactual notation to define causal effects:**
- *Individual-level causal effect:*
  - $Y_{a=1} - Y_{a=0}$
  - $Y_{a=1} / Y_{a=0}$
    - **Note:** this is inherently unobservable. Nor can we use data to estimate this in a meaningful way.
- *Population-level causal effect:*
  - $E(Y_{a=1}) - E(Y_{a=0})$
  - $E(Y_{a=1}) / E(Y_{a=0})$
    - **Note:** this is inherently unobservable, but we can use data to approximate this counterfactual quantity.

Counterfactual and observed data

- *Causal contrast (unobservable):*
  - $E(Y_{a=1}) - E(Y_{a=0})$
  - $E(Y_{a=1}) / E(Y_{a=0})$
- *Real-world contrasts (observable):*
  - $E(Y|A=1) - E(Y|A=0)$
  - $E(Y|A=1) / E(Y|A=0)$
- **The goal of counterfactual causal inference** is to use real data and methods to approximate causal quantities:
  - *We want to use:* $E(Y|A=1) - E(Y|A=0)$

    to approximate: $E(Y_{a=1}) - E(Y_{a=0})$

Causal inference as a missing data problem

Approximating the counterfactual

- **Assumptions:**
  - To what degree do our real, observed data support the process of inferring causal effects?
  - Exchangeability, positivity, consistency.
- **Methods** for using real data to approximate counterfactual quantities:
  - Inverse probability weighting (IPW)
  - G-computation and g-estimation

1. Positivity

"Every participant has some **positive** probability of receiving each treatment setting."

- Also known as "common support"
- If some group cannot receive treatment, then we can't estimate a causal effect among this group.
- If confounders *deterministically assign* exposure among some group, causal inference is not possible.

$$\Pr(A=1 \mid \overline{W} = \overline{w}) = 1.0$$

$$\Pr(A=1 \mid \overline{W} = \overline{w}) = 0.0$$

Positivity

- ***Consider the question:*** *What is the effect of epidural analgesia on PPH?*
- Say we conduct an RCT with 2 groups: epidural versus no anesthesia.
  - In this RCT, all women have a known, positive probability of receiving epidural (exposed) and no anesthesia (unexposed):
    - Pr(epidural)=0.5
    - Pr(no anesthesia)=0.5
  - *Positivity assumption is met by design.*
- This is not guaranteed in the context of observational data.

Non-positivity

- Pr(epidural)=0:
  - Women delivering at home or free-standing birth centers.
  - Epidural analgesia is not available in these settings.
- Pr(no anesthesia)=0:
  - Women delivering by scheduled cesarean delivery
  - All will have some form of anesthesia (e.g., epidural, spinal, general)
- These are violations of the positivity assumption: these groups have a *non-positive probability* of receiving a treatment setting.

Non-positivity

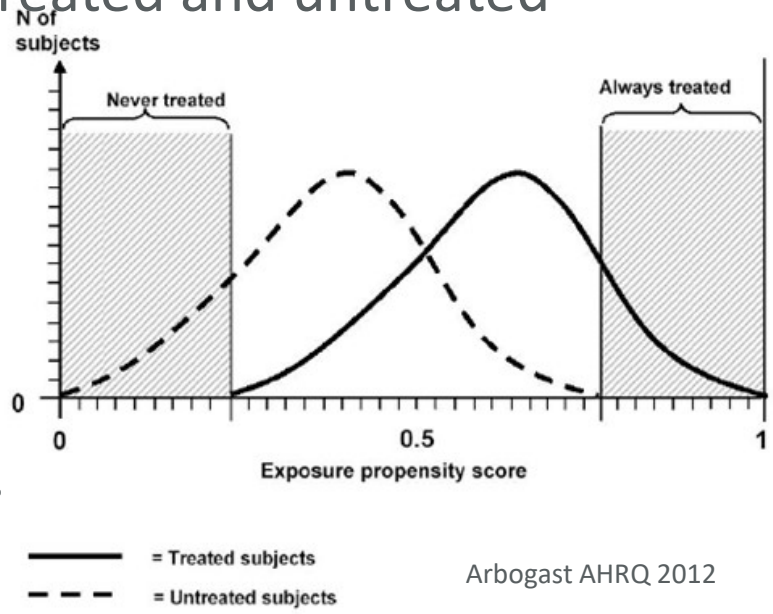
- Violations of the positivity assumption are also called “Lack of common support”
  - More common in the social sciences (e.g., Manski 1993)
- If someone cannot possibly receive a treatment, how we infer causal effects of the treatment in this person?
  - It may be computationally possible, i.e., a model will converge.
  - But causal interpretation is debatable or untenable.

Non-positivity

- Deterministic:
  - Individuals in ≥1 data stratum cannot receive a given exposure setting.
    - "Data stratum" is defined as stratum of the confounders.
  - Epidural non-positivity examples (scheduled cesarean and home birth) are *deterministic*.
  - You could not collect more data and address the issue.
- Random:
  - Individuals in ≥1 data stratum did not receive a given exposure setting, e.g., by bad luck.
  - Say you did a small study (N=300) study of laboring women, and some stratum (e.g., white women >35 y.o. with private insurance) happened to have Pr(epidural)=1.
  - This is a *random* positivity violation: if you enrolled more participants, you would eventually have both exposure settings in this stratum.

Non-positivity

- Especially complex for continuous exposures.
- Techniques to assess for positivity assumption violations:
  - **Stratified analysis:** within each strata defined by confounders, are there both treated and untreated observations?
  - **Multivariable:** use the propensity score to condense confounders into 1 variable, then examine overlap between treated and untreated by the PS.



Arbogast AHRQ 2012

2. Exchangeability

"After accounting for all measured confounders, the only difference between the exposed and unexposed groups is their exposure status (ie, they are otherwise exchangeable)."

- No unmeasured confounding.

Exchangeability

Among US Service members deployed in the

Global War on Terror, does blast exposure cause perceptual hearing deficits?

Population: US Service members

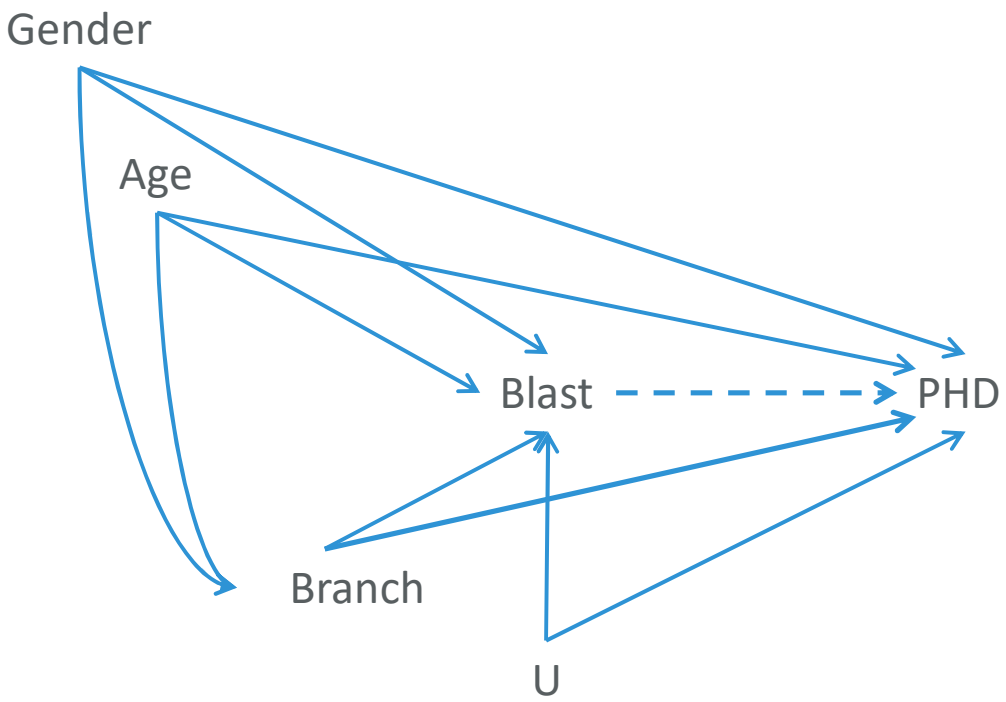
Etiologic Time Frame: 10/2001 - current

Exposure: Blast

Outcome: Perceptual hearing deficits (PHD)

Idealized RCT: Randomly select Service members, then randomize blast exposure

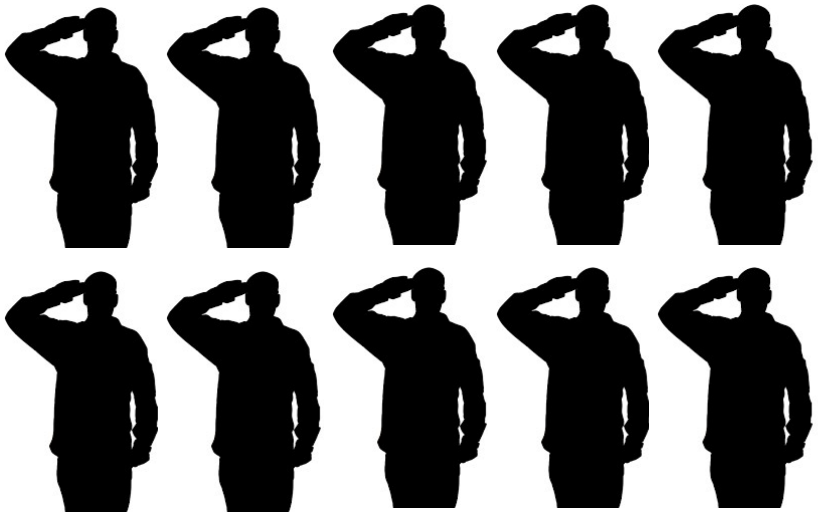
Exchangeability



Exchangeability

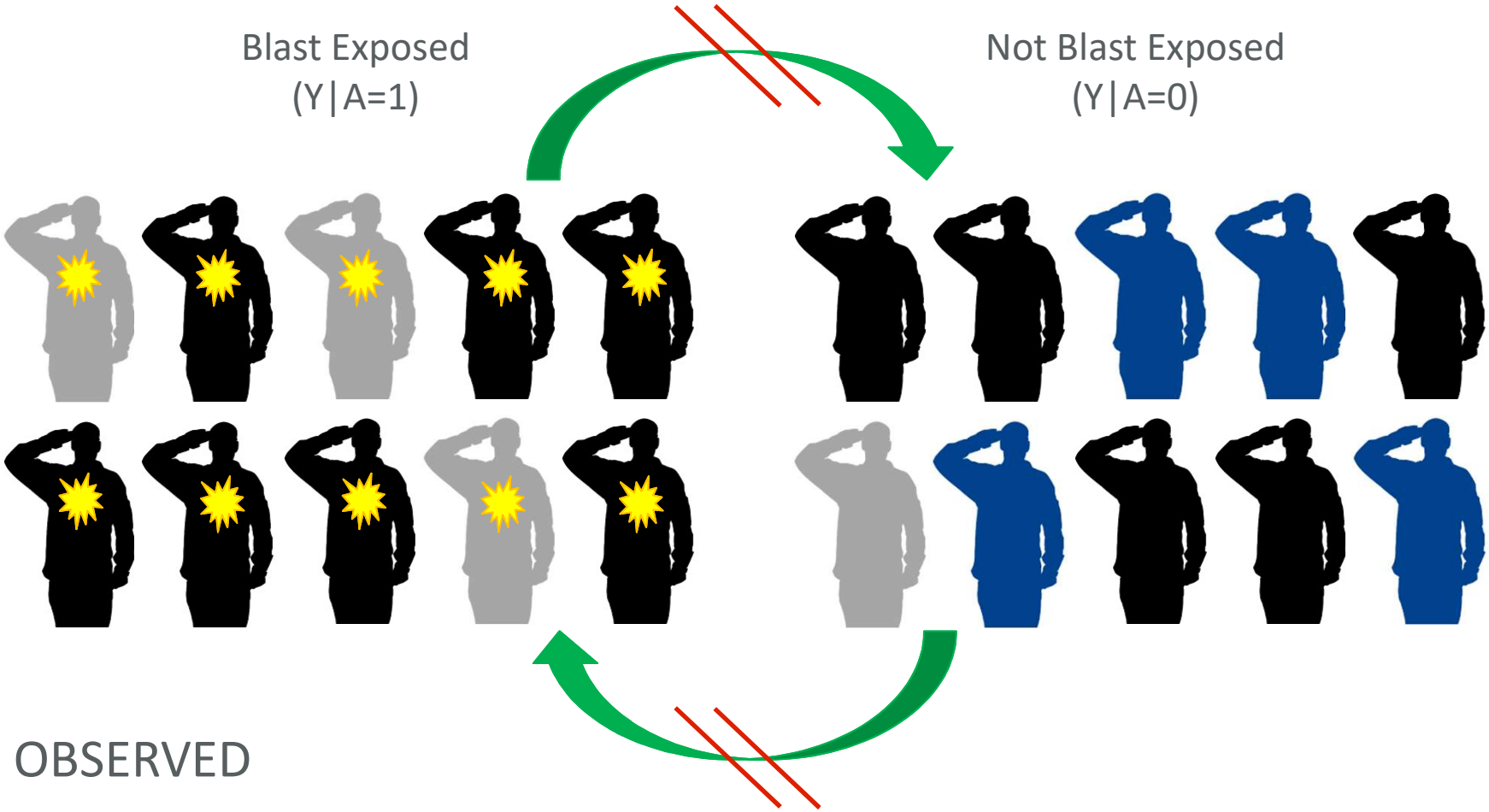
Blast Exposed
($Y_{a=1}$)

Not Blast Exposed
($Y_{a=0}$)



COUNTERFACTUAL

Exchangeability



OBSERVED

Exchangeability

- Ways to achieve “no unmeasured confounding”:
  1) No confounding: Successful randomization, with no post-baseline confounding.
  2) If confounding present: reliably collect data on all confounders, adjust for them all (e.g., as in a correctly-specified model)
- This is a theoretical ideal rather than something we can expect in reality (let alone guarantee).
- It requires that we prevent confounding at the design phase (e.g., RCT), or correct for it (using deep subject matter knowledge, accurate data for all variables, and correct model specification).

Exchangeability

- Even with the same driving causal questions, potential confounders may not be the same in every study
- Untestable assumption
  - Conduct bias analysis to determine impact of an unmeasured confounder on observed association

RCTs and causal assumptions

- In RCTs, data are generated for the purpose of addressing the study question.
- Therefore, a strength of RCTs is that these causal assumptions are met by design:
  - All enrolled participants have a positive probability of receiving all treatment settings.
  - Asymptotically, exchangeability will be achieved (at baseline).

3. Consistency

"One needs to be able to explain how a certain level of exposure could be hypothetically assigned to a person exposed to a different level" S. Cole, 2009

- This is a bit philosophical.
- It invokes several hypothetical concepts:
  A. The existence of counterfactuals.
  B. The hypothetical ability to intervene and alter observed exposure to another (unobserved) setting.
  C. The unobserved, new exposure causally acting on outcome, producing an unobserved, counterfactual outcome.
  D. The concordance of this counterfactual outcome with something we'd observe in the real world.

Consistency

- How are hypothetical exposure changes achieved?
- For example: how could someone go from BMI 45 to 35?
  - Diet? Weight loss? Bariatric surgery?
  - Are these changes contained in your data?
  - Would all of these changes result in the same effect?

    (*treatment version irrelevance*)

	Exposure scenarios	
Participant #	**1. Observed BMI** (kg/m^2)	**2. Counterfactual BMI, all values of BMI>35 decreased to 35**
…	…	
5,000	45	**35**

Assumptions: theoretically

- If positivity, exchangeability and consistency truly held, then our calculated measure of association equals the causal association:

$$E(Y|A=1) - E(Y|A=0)$$

$$=$$

$$E(Y_{a=1}) - E(Y_{a=0})$$

Agenda

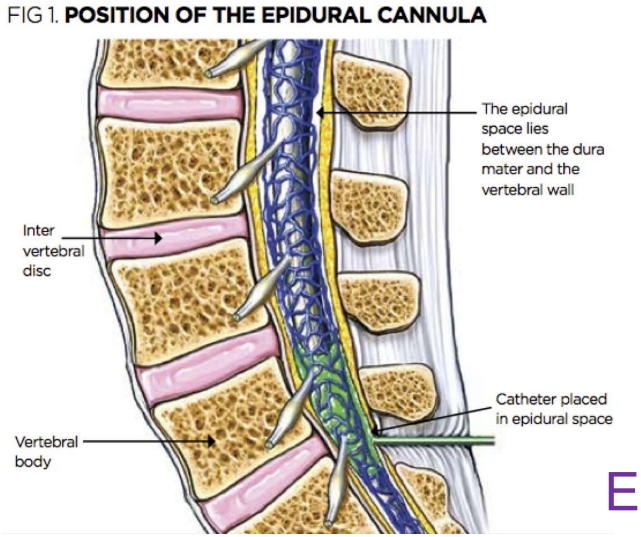
- Defining CI, avoiding misconceptions
- Notation
- Assumptions
- 3-step process:
  - Step 1: Formulate question
  - Step 2: Assess data
  - Step 3: Design analysis
- Break

Step 1: Formulate the question

- Everyone agrees that good questions are important.
  - Perhaps even more so for CI.
- ***How do we define this concept?***
- This is the first stage of CI, from which the other steps follow.
- Recommend against letting your question be driven by factors such as outcome or exposure coding.

Ideal experiment framework

- What is the effect of epidural analgesia on 2nd stage labor duration?



Epidural analgesia → Labor duration

Ideal experiment framework

- What is the effect of epidural analgesia on 2nd stage labor duration?
- What if we conducted an RCT of epidural analgesia?
  - Sure.
  - People have.

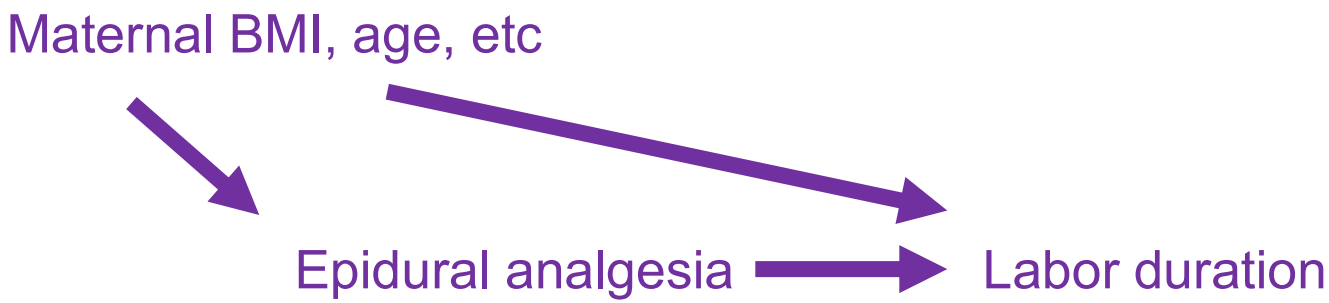
Ideal experiment framework

- ***Using secondary data,*** what is the effect of epidural analgesia on 2nd stage labor duration?

Epidural analgesia → Labor duration

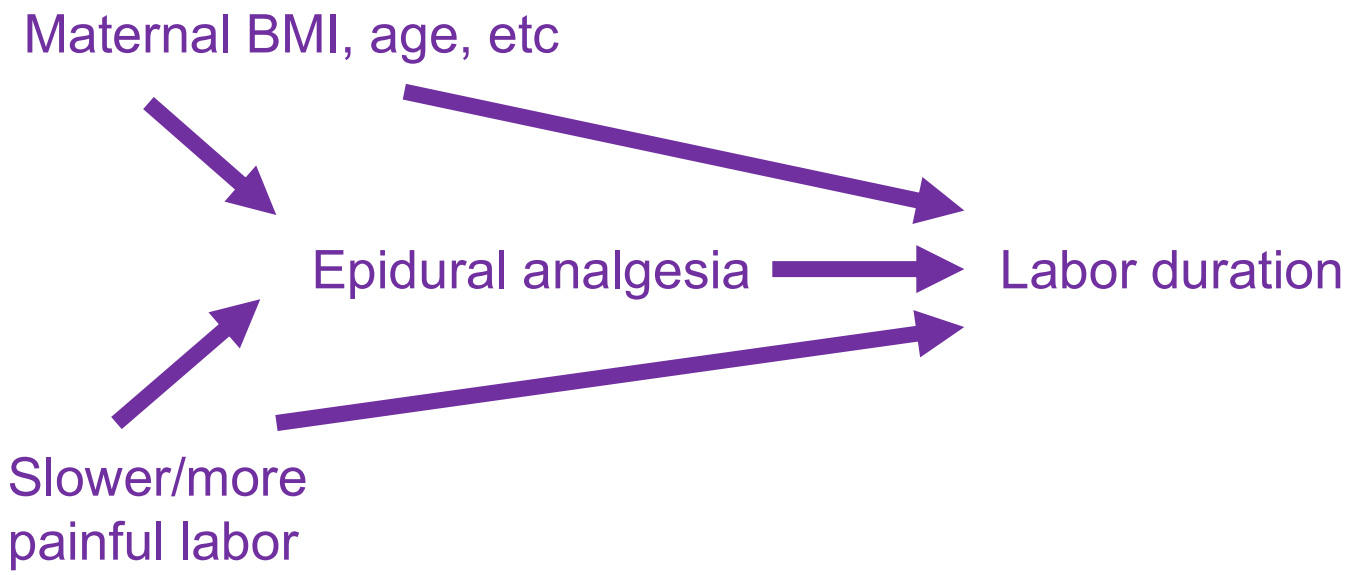
Ideal experiment framework

- ***Using secondary data,*** what is the effect of epidural analgesia on 2nd stage labor duration?
  - Controlling for confounding.



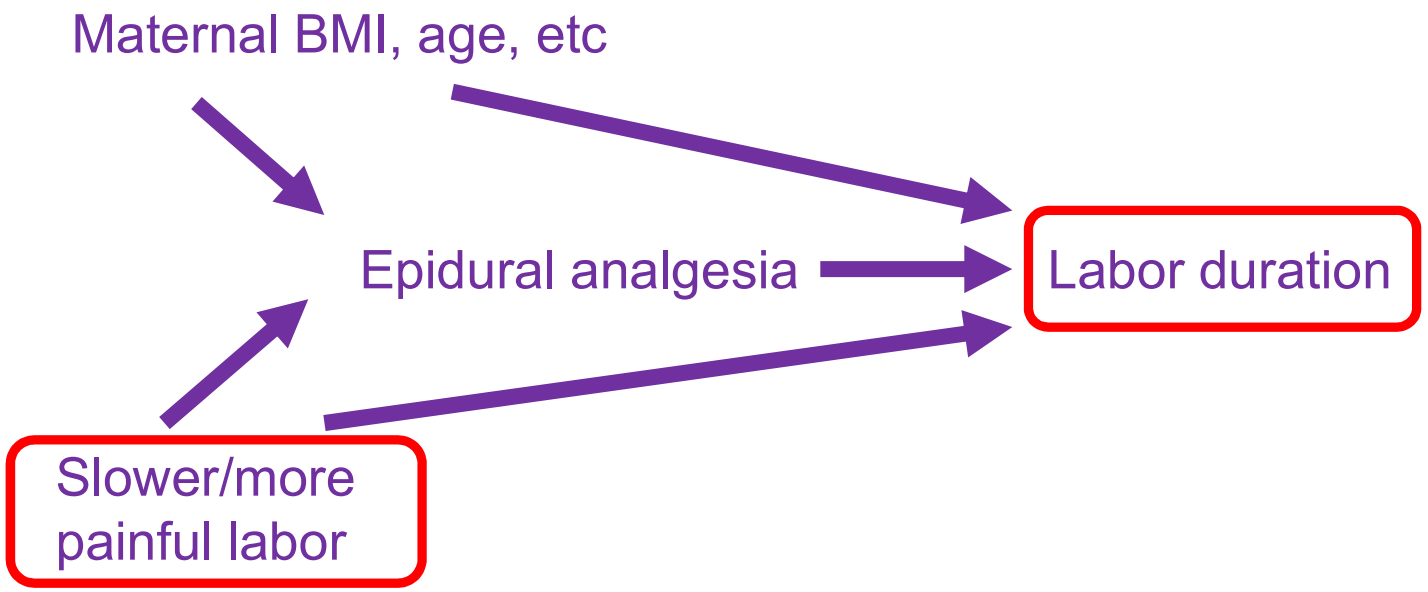
Ideal experiment framework

- ***Using secondary data,*** what is the effect of epidural analgesia on 2nd stage labor duration?
  - Controlling for confounding, including by labor duration.
  - Slower, more painful labor → epidural.



Ideal experiment framework

- Note that labor duration predicts the exposure, and is the outcome.
- Confounding by indication.
- How do we address this sticking point?

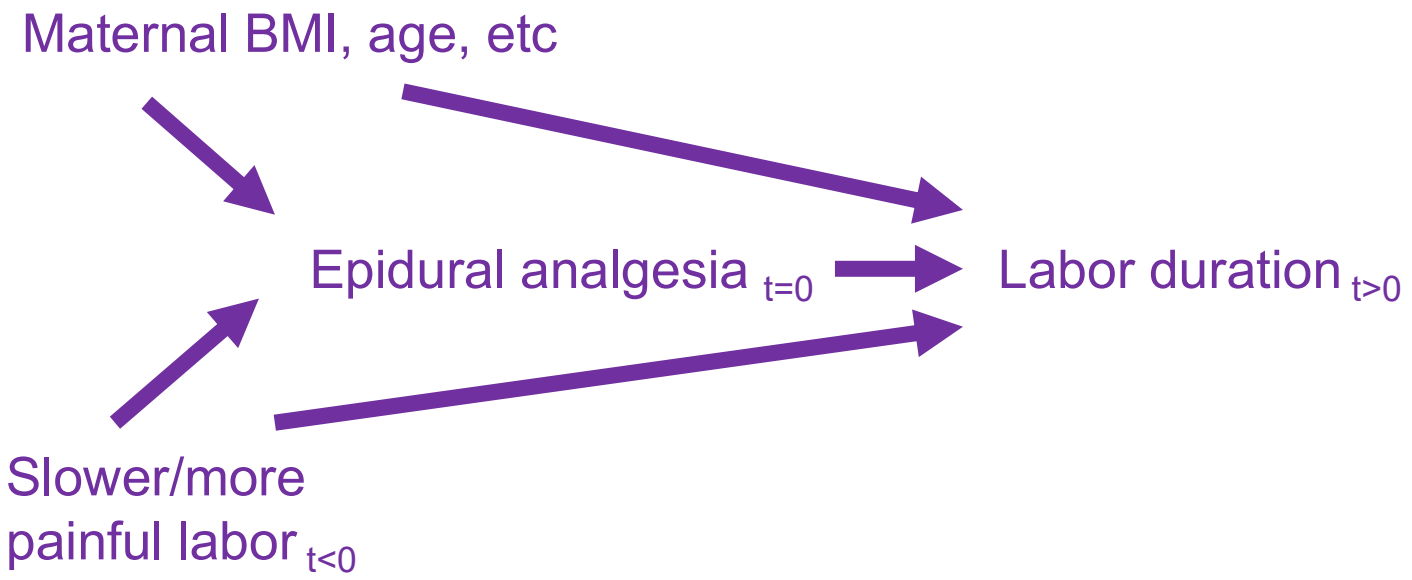


Ideal experiment framework

- We address this by more precisely defining our exposure and outcome, and in turn confounders.
- **Exposure:** epidural administered at time ***(t=0)***
- **Outcome:** 2nd stage duration, after time t=0 ***(i.e., t>0)***

Ideal experiment framework

- We address this by more precisely defining our exposure and outcome, and in turn confounders.
- **Confounders:**
  - **Baseline:** maternal BMI, age, race, etc. These do not change in the study time of interest.
  - **Time-varying:** Labor duration before time t=0 ***(i.e., t<0)***



Ideal experiment framework

- We have more precisely defined our causal question invoking the concept of temporality.
- There are still complexities:
  - We have anchored exposure timing ***(t=0)*** for exposed women, but not unexposed women (who did not receive epidural)
  - We have defined timing of exposure initiation, but not dose, frequency, etc.
    - Do these matter here?
  - We have defined the data needed to ask this precise question: but do we have such data?
    - Data on timing of events in the labor course are extremely rare.
- We have raised more questions, and have started down the path of defining a sound causal question.

Agenda

- Defining CI, avoiding misconceptions
- Formulating answerable questions
  - CI in samples of infants born preterm
- 3-step process:
  - Step 1: Formulate question
  - Step 2: Assess data
  - Step 3: Design analysis
- Break

Step 2: Assess the data

- Next, consider the fit between the causal question and the data source.
- Do the available data permit assessment of the causal question?
  - Are the necessary variables present?
  - Are they measured well?
  - Is there sufficient granularity?
  - Is there empirical data support?

Step 2: Causal assumptions

- In RCTs, data are generated for the purpose of addressing the study question.
- Therefore, a strength of RCTs is that some assumptions are met by design.
- ***Exchangeability*** (asymptotically)
- ***Positivity***
  - every participant has some **positive** probability of receiving each treatment setting.
  - Treatment settings: (1) epidural; (2) no anesthesia
- ***Consistency***

Step 2: Assess the data

- Questions in assessing the data:
  - *Are there positivity violations?*
  - *Do the data contain the necessary variables and granularity?*
    - E.g., is there timing of epidural administration?
    - If so, index exposure, outcome, and confounders.
    - If not:
      - Use another data source.
      - Reframe your question so it is not one that is explicitly causal in nature.

Paediatric and Perinatal Epidemiology

doi: 10.1111/ppe.12387

Maternal Body Mass Index and Regional Anaesthesia Use at Term: Prevalence and Complications

Frances M. Biel,[a] Nicole E. Marshall,[a] Jonathan M. Snowden[a,b]

- What are the associations between maternal BMI and: (1) epidural analgesia; (2) cesarean delivery; (3) epidural complications?

A

Maternal BMI → Epidural analgesia → Cesarean delivery

B

Maternal BMI → Cesarean delivery → Epidural anesthesia

Agenda

- Defining CI, avoiding misconceptions
- Formulating answerable questions
  - CI in samples of infants born preterm
- 3-step process:
  - Step 1: Formulate question
  - Step 2: Assess data
  - Step 3: Design analysis
- Break

Step 3: Design an analysis plan

- Carry forward knowledge from steps 1 and 2:
  - Sharp causal contrasts:
    - Define all settings of the treatment variable in a clear and relevant way.
    - Respect the temporality criterion.
  - **Positivity assumption:**
    - Restrict analysis to areas of common support
    - Or, choose an estimator that does this (e.g., PS matching)

Step 3: Design an analysis plan

- **Exchangeability assumption:**
  - Almost never met in observational data
  - Therefore, use your analytical tools to address:
    - Restriction, multivariable analysis, stratification…

Analysis: Choose tools wisely

- Consider special features of your data.
- ***Strengths:***
  - Instrumental variables
  - Policy change

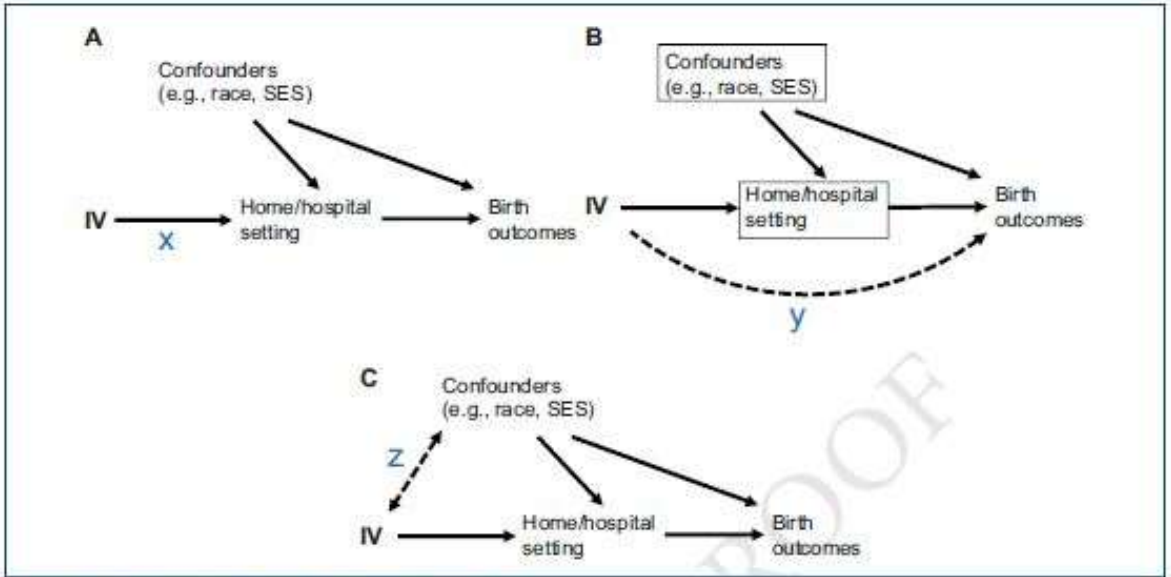


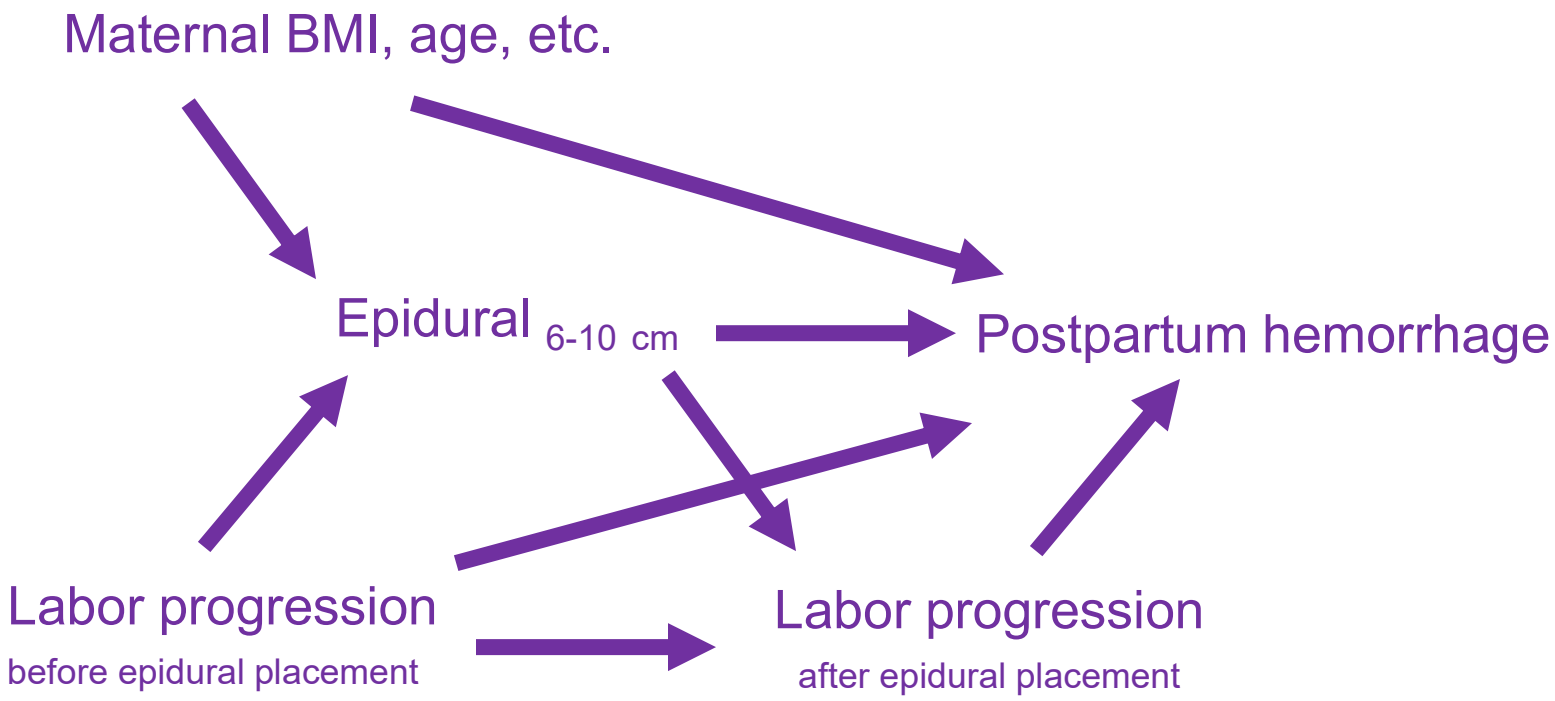
Figure 2. Illustration of Assumptions Required of IV Analysis
A) Assumption 1: Instrumental variable affects exposure. B) Assumption 2: Instrumental variable is not associated with outcome, except through exposure. C) Assumption 3: Instrumental variable is not associated with confounders.
Abbreviations: IV, Instrumental variable; SES, socioeconomic status.

Analysis: Choose tools wisely

- Consider special features of your data.
- ***Challenges:***
  - Say we want to estimate the total effect of epidural analgesia on PPH.
    - **Total effect:** effect of exposure on outcome, flowing through all causal pathways/mechanisms.

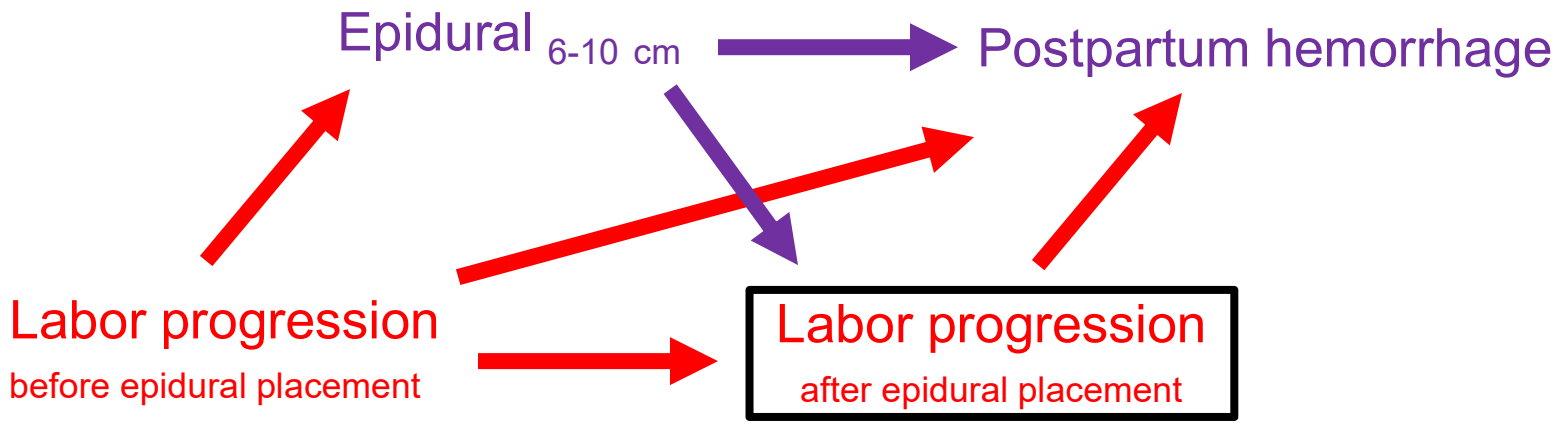
Analysis: Choose tools wisely

- Consider special features of your data.
- ***Challenges:***
  - Time-dependent confounding (epidural → PPH)



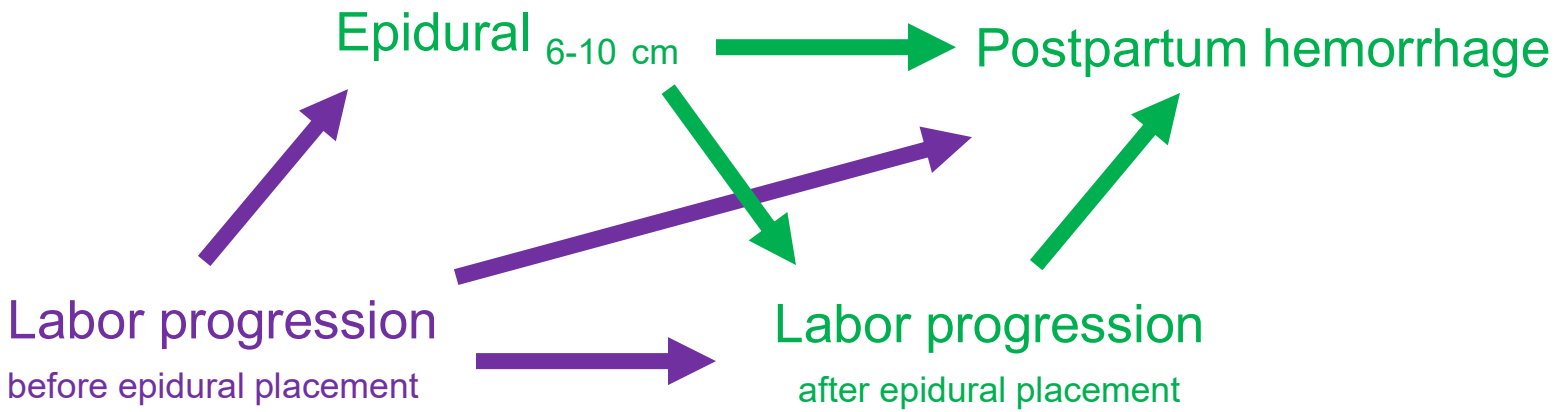
Analysis: Choose tools wisely

- Consider special features of your data.
- ***Challenges:***
  - Time-dependent confounding
  - **Confounding pathway**
    - Suggests controlling for intermediate



Analysis: Choose tools wisely

- Consider special features of your data.
- ***Challenges:***
  - Time-dependent confounding
  - **Causal pathway**
    - Suggests against controlling for intermediate

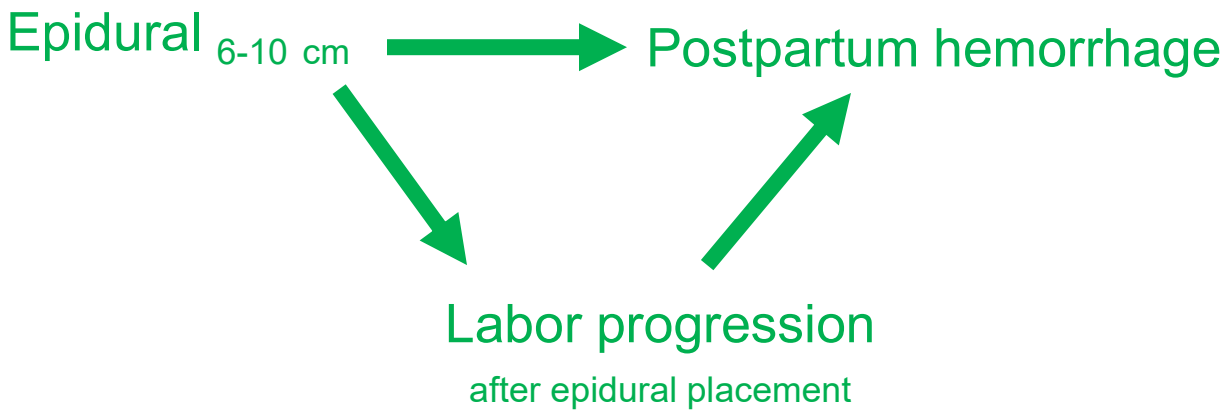


Analysis: Choose tools wisely

- In this case of time-dependent confounding, g-methods enable confounder control without conditioning on a causal intermediate.
  - IPW
  - G-computation
  - G-estimation
  - Doubly robust methods

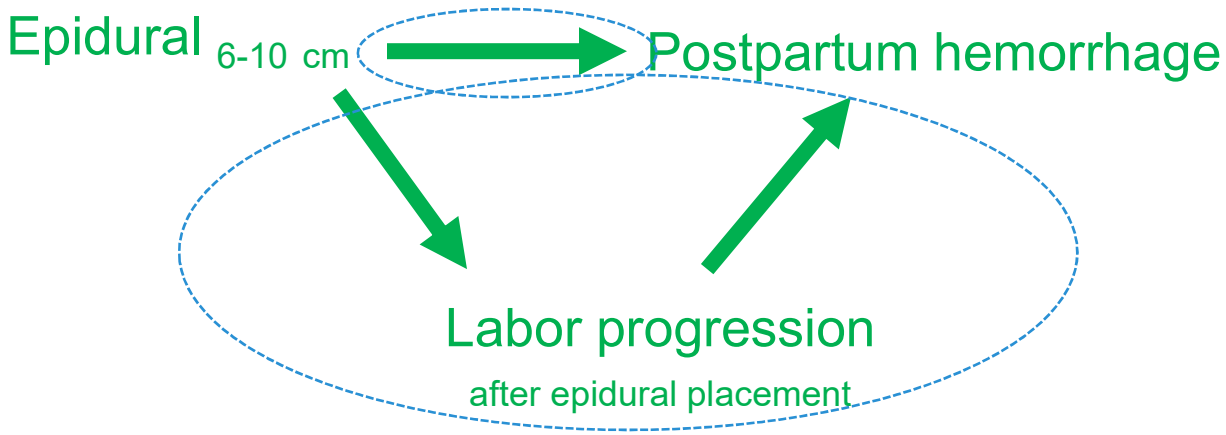
One question/tool may lead to a different, new question.

- **What is the mechanism of epidural effects on PPH?**
  - This question contrasts with the total effects analysis we targeted before, despite coming from the same DAG.
  - Here, we seek to quantify the mechanism that explains ***how*** epidural use affects PPH.
    - How much of the effect is explained by longer labor (post-epidural), versus other mechanisms (e.g., contraction strength, maternal positioning, etc).



One question/tool may lead to a different, new question.

- **What is the mechanism of epidural effects on PPH?**
  - Formally:
    - What is the ***direct effect*** of epidural use on PPH?
    - What proportion of the effect is mediated through labor progression (i.e., ***indirect effect***)?
  - This question concerns mediation (effect decomposition), which we discuss later.



Break

When we return:

Selecting and framing a maternal health question

II: Selecting and framing a maternal health question

Agenda

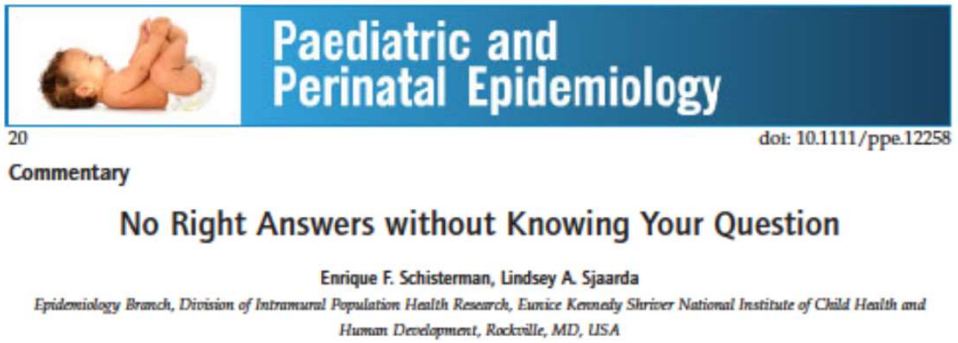
- Question/methods decision tree
- Formulating a question:
  - The effects of obesity on PPH.

Agenda

- Question/methods decision tree
- Formulating a question:
    - The effects of obesity on PPH.

Leading with the question

- Rigorous question formulation is essential for sound epidemiologic science.



- However, we lack consensus on what constitutes a "good question," and how you formulate one.
- We need approaches to guide question formulation and selection of corresponding methods.

Leading with the question

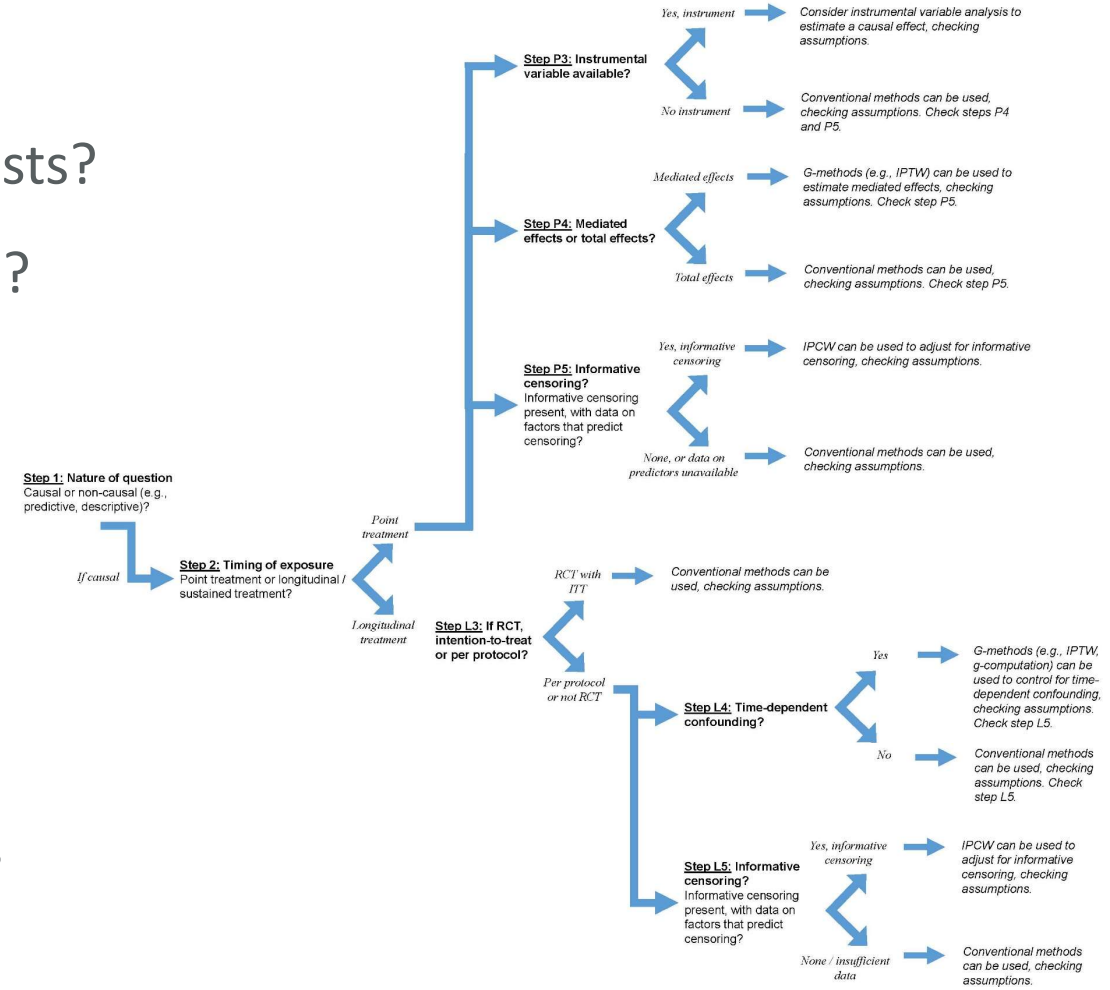
- We propose 1 approach to question formulation and selection of methods.
- Features of your question and data suggest analytical approaches.
  - Is exposure defined at one time? Or is exposure a time-varying process?
  - Is there an instrumental variable available?
  - Are total effects or mediated effects of interest?
- We propose a branching decision tree to help determine which method is called for ("indicated").
  - Based loosely on the CERBOT (http://cerbot.org/) system, by Yi Zhang *et al*.

Question/methods decision tree

- Having formulated a causal question, enables the investigator to identify which features of the question are relevant to selecting a methodological approach.
  - "The question" includes the study/data used to operationalize the given research question.
- The presence/absence of each feature suggests a given method.
- Some methods may be used in combination.
- This process is iterative: considering the features of your question (and data available to answer it) may prompt you to re-tool your causal question.
- ***This is merely 1 proposed approach, highlighting counterfactual CI.***
- ***There are other factors to consider and there is no "cookbook."***

Question/methods decision tree

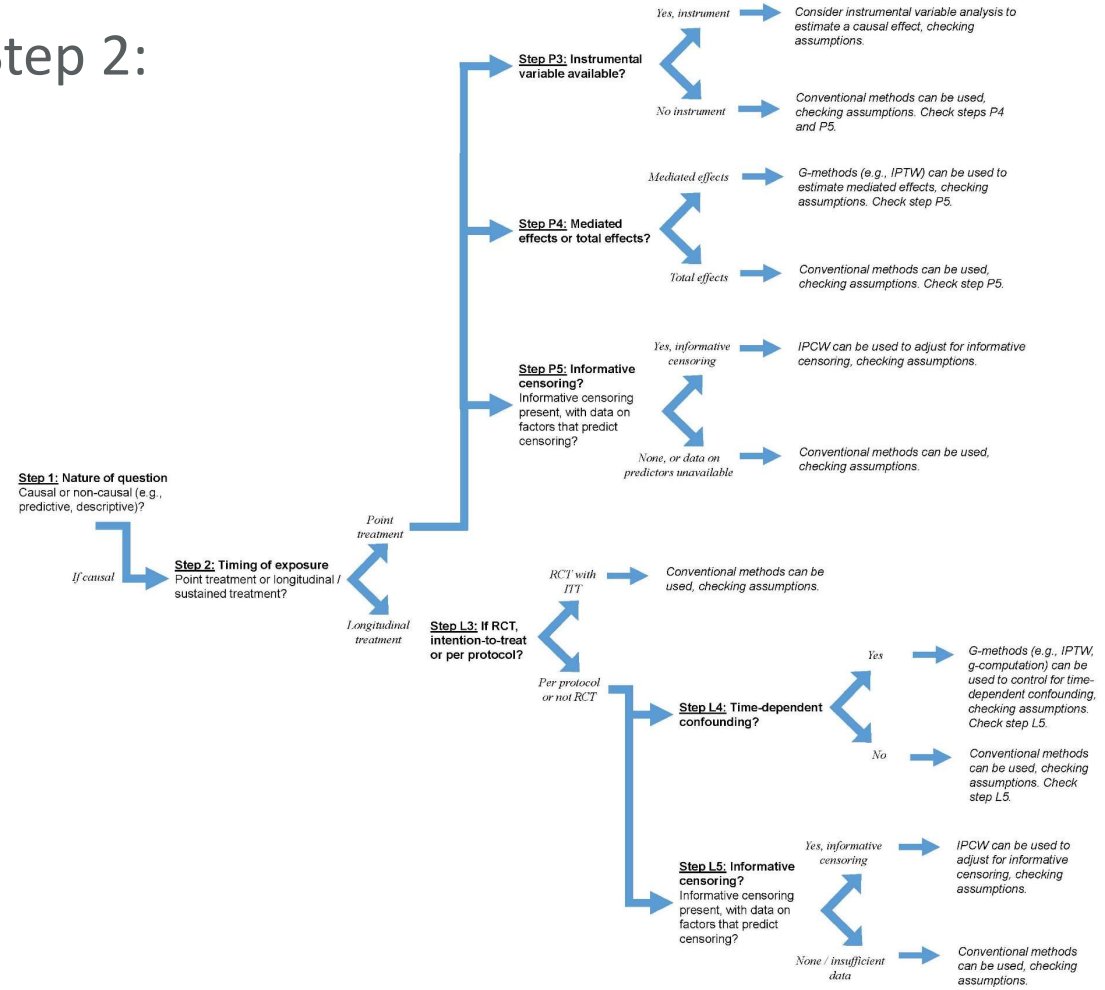
- Questions include:
  - exposure timing?
  - instrumental variable exists?
  - total or mediated effects?



- This is a bit complicated.

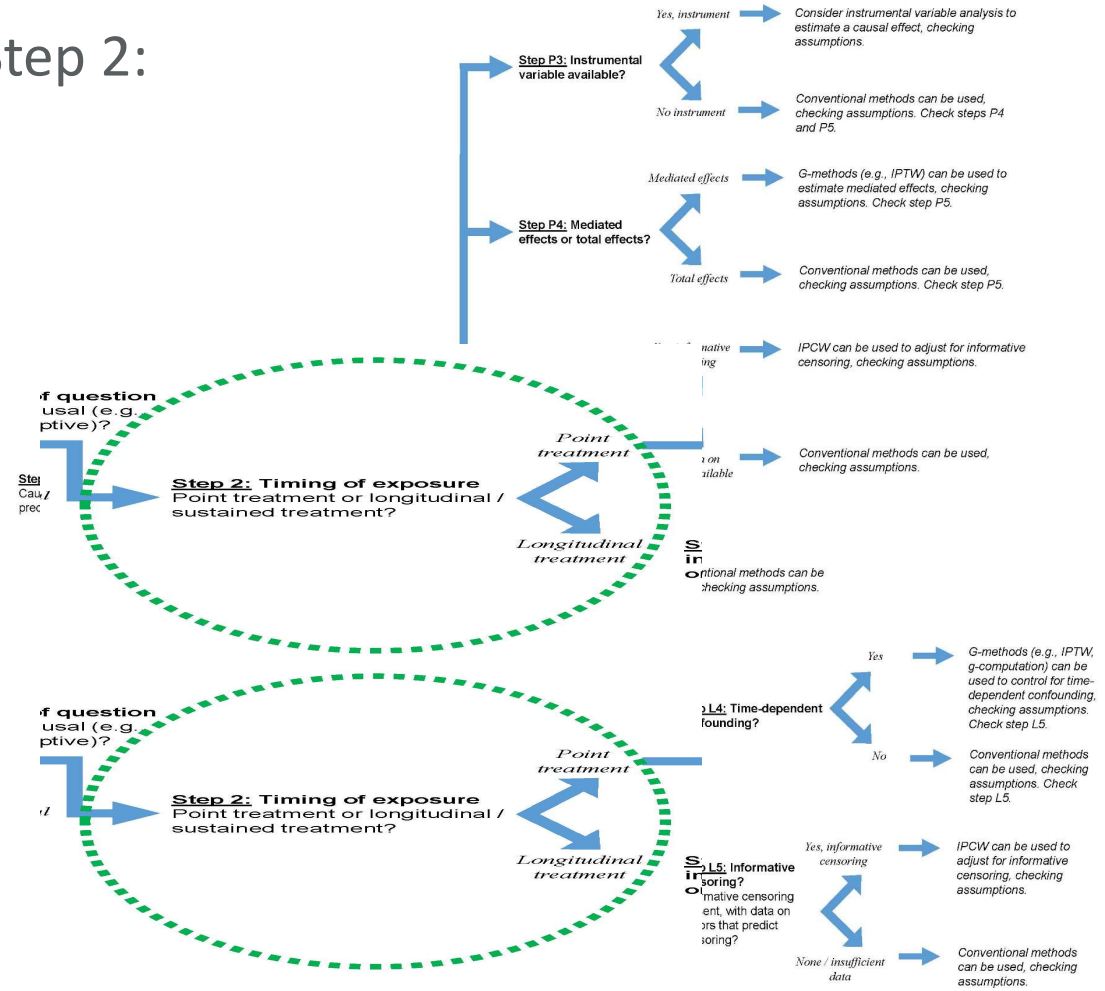
Question/methods decision tree

- To simplify:
  - We suggest breaking at Step 2:



Question/methods decision tree

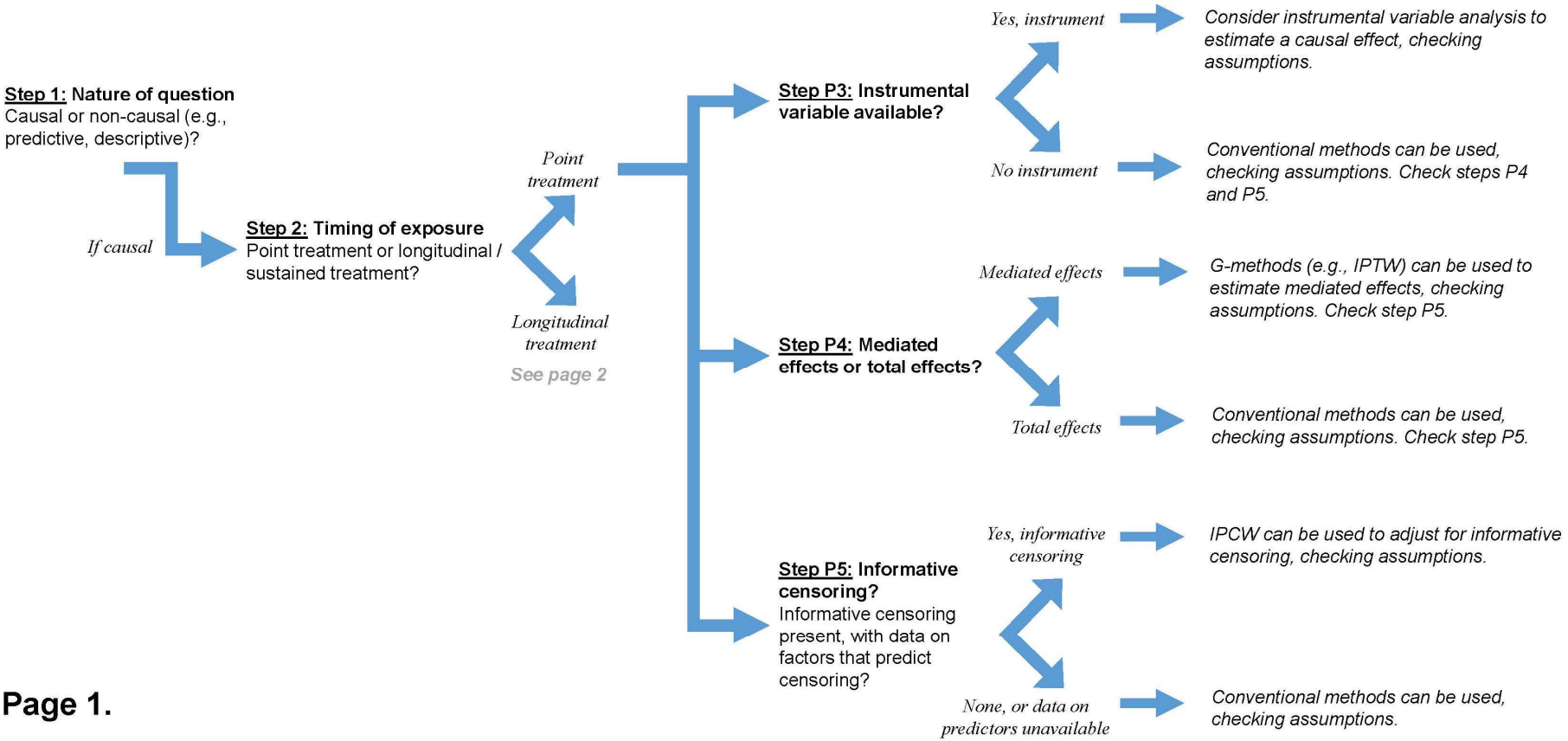
- To simplify:
  - We suggest breaking at Step 2:



*Is treatment a **point treatment** (measured/occurs at one point in time) or a **longitudinal/sustained treatment** (measured/occurs over time, and may change)?*

Question/methods decision tree

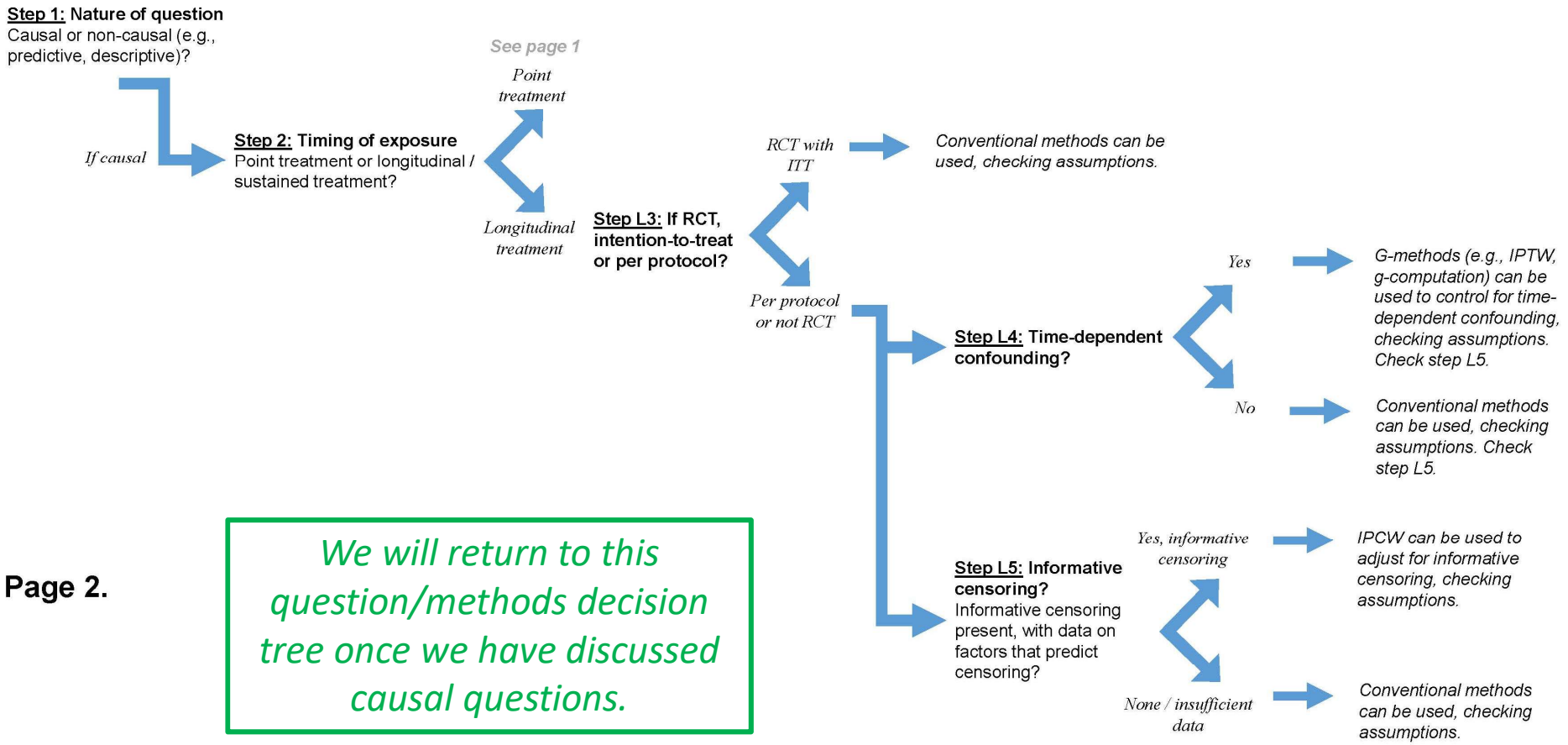
- Point treatment.



Page 1.

Question/methods decision tree

- Longitudinal/sustained treatment.



Page 2.

We will return to this question/methods decision tree once we have discussed causal questions.

Agenda

- Question/methods decision tree
- Formulating a question:
  - The effects of obesity on PPH.

Maternal health in the US

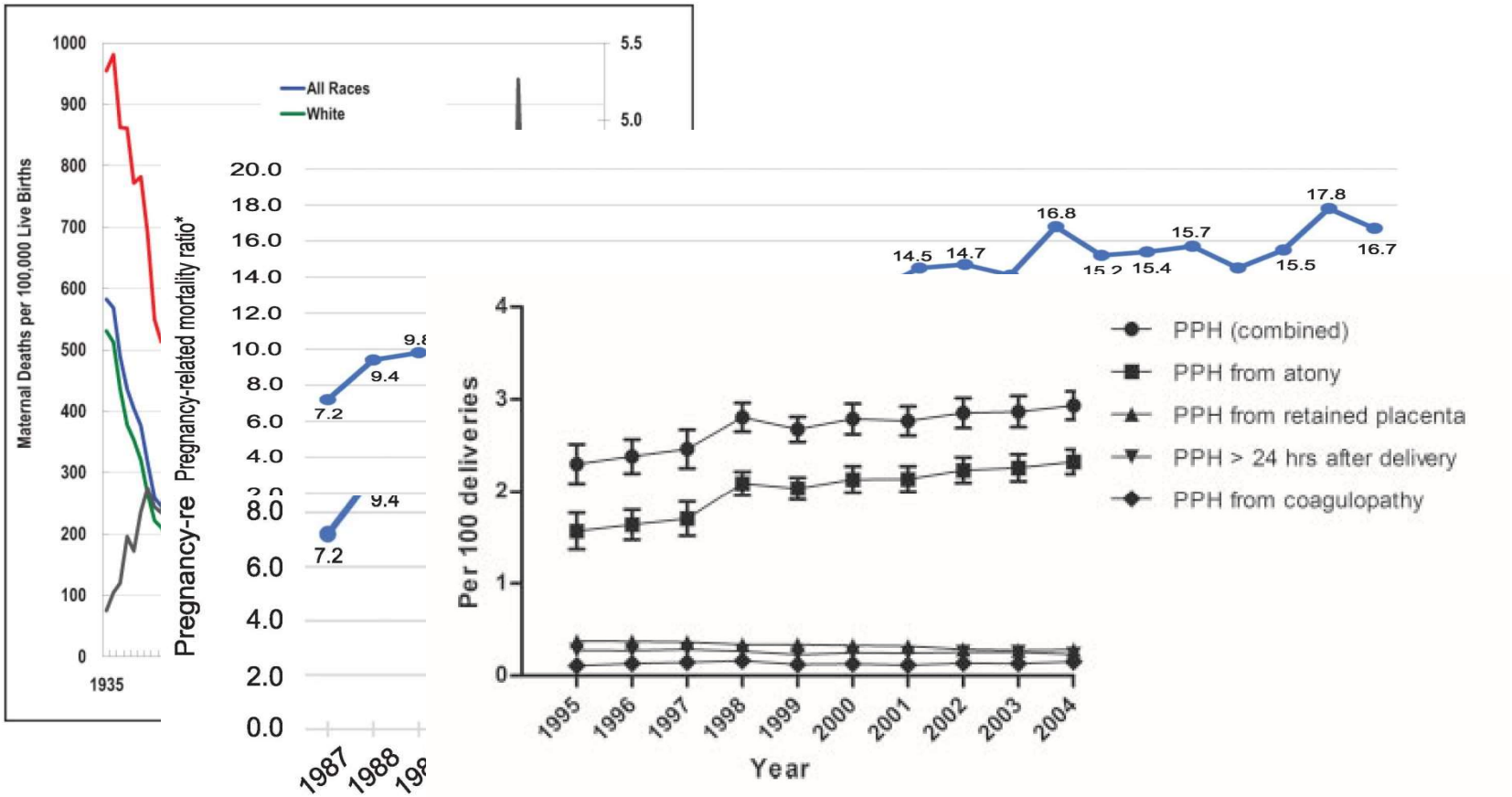


Figure 1. Trends in the rate of overall postpartum hemorrhage and hemorrhage by underlying etiology: 1995 to 2004. Annual rates and 95% confidence intervals are displayed.

Bateman, 2010

Meeting the challenge to improve maternal health

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

NOVEMBER 1, 2018

What
— Al
Susan M

Postpartum
for a New Pa

Mara E. Murray Horwit

The American College of Obstetricians and Gynecologists
WOMEN'S HEALTH CARE PHYSICIANS

COMMITTEE OPINION

Number 666 • June 2016

Optimizing Postpartum Care

Committee on Obstetric Practice

Maternal health research questions

- What is the unit of analysis?
  - Mother/birthing person, as compared to the infant
- How we manage and analyze our data are influenced by this.
  - Multi-fetal gestations are only one observation (i.e., row)
  - Different exclusions (e.g., include placenta previa, preterm birth)
  - Women with prior cesarean are included in the main analytical sample.
- Where do variables fall in the causal pathway?
  - Maternal characteristics and pre-pregnancy morbidities
  - Pregnancy-related conditions
  - Intrapartum/childbirth care
- What are important questions in maternal health?
  - Here, let's focus on severe maternal morbidity (SMM)

Predictors of SMM

- Maternal age
- Maternal race/ethnicity (i.e., structural racism, health care access, etc)
- Maternal body mass index (BMI)
- Health care access and utilization
- Pre-pregnancy morbidities (e.g., chronic hypertension [htn.])
- Pregnancy-related morbidities (e.g., GDM)
- Quality of care (prenatal, intrapartum)
- Childbirth care including procedure use

Predictors of SMM

- We know many individual predictors, but how do they all fit together?
  - I.e., what is their causal structure?
- What are the "targets" that we could intervene upon, to prevent SMM?
- Obesity is a persistent, substantial predictor of SMM.
  - What about obesity explains this increased risk?
  - What are the mechanisms?

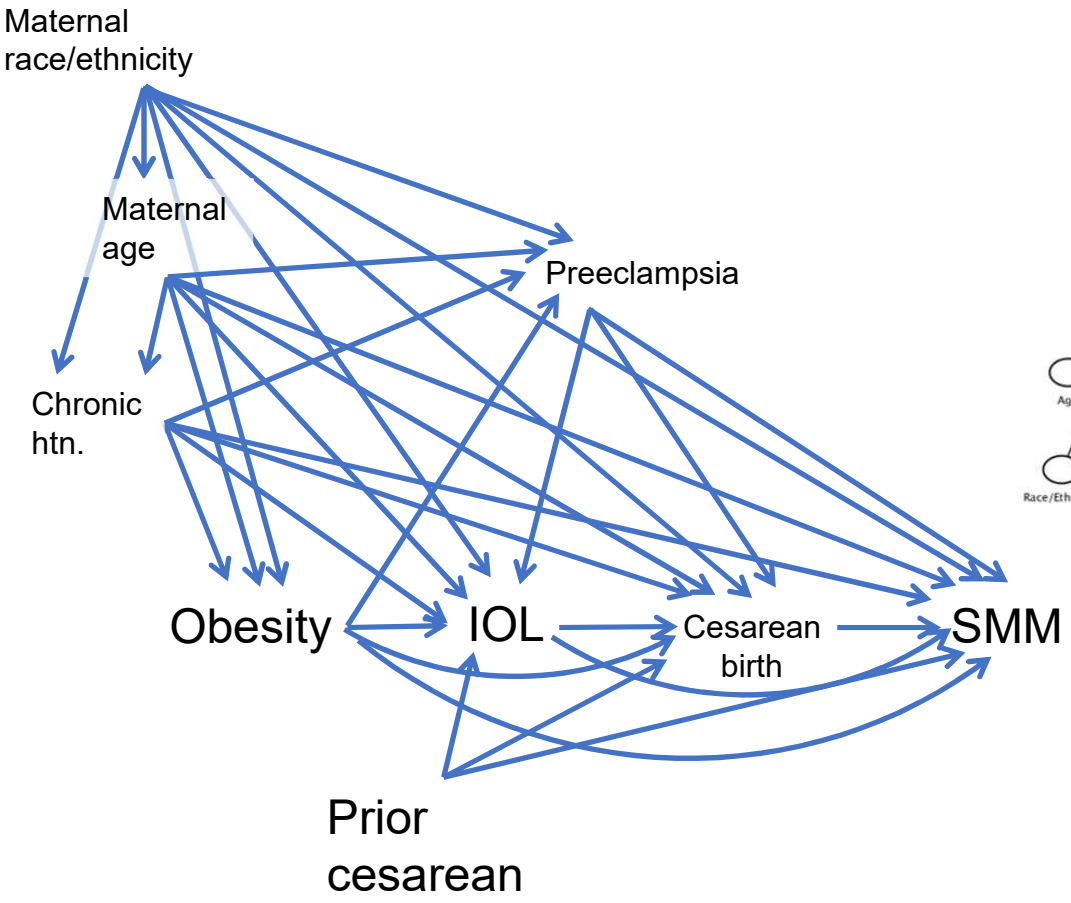
Activity: causal diagrams

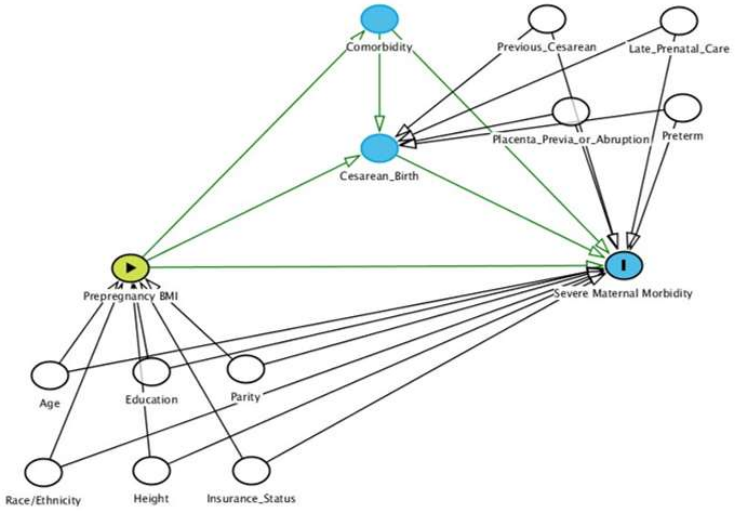
- Causal structure of the obesity/SMM association.
- Using these variables, construct a causal diagram depicting obesity (exposure), SMM (outcome), and other relevant variables:

  - ***Obesity (exposure)***
  - ***SMM (outcome)***
  - Maternal age
  - Chronic (pre-pregnancy) htn.
  - Cesarean birth
  - Preeclampsia
  - Induction of labor
  - Maternal race/ethnicity
  - Prior cesarean

Our posited causal diagram

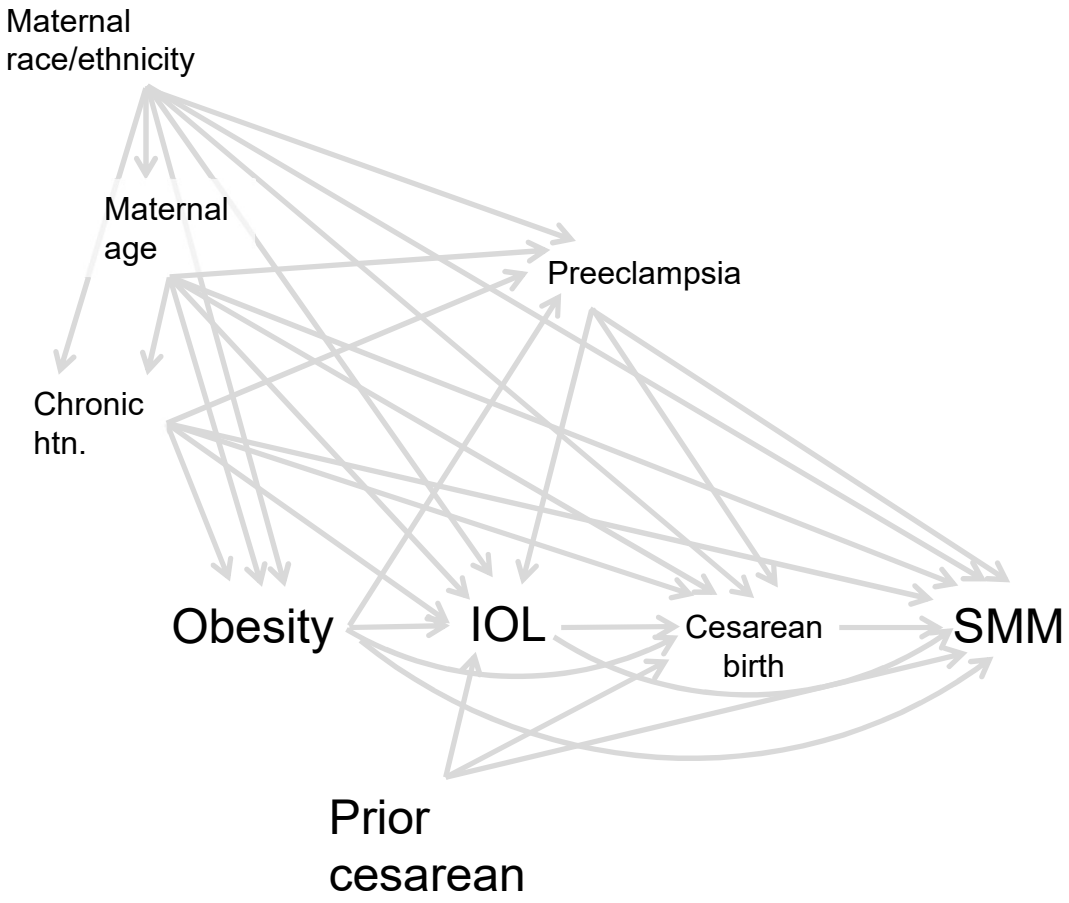
Leonard et al. *Risk of Severe Maternal Morbidity in Relation to Prepregnancy Body Mass Index: Roles of Maternal Comorbidities and Cesarean Birth*



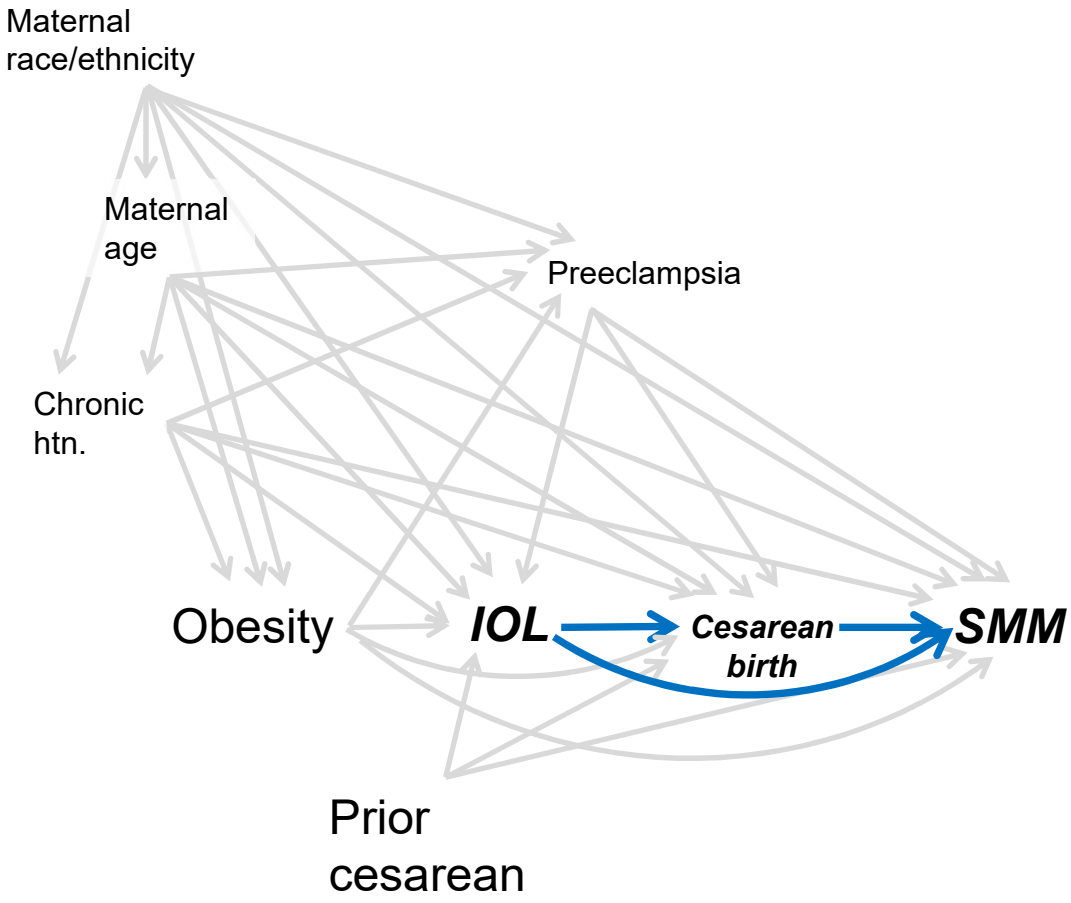


*The causal structure of many research questions is **complicated.***

One DAG, many questions

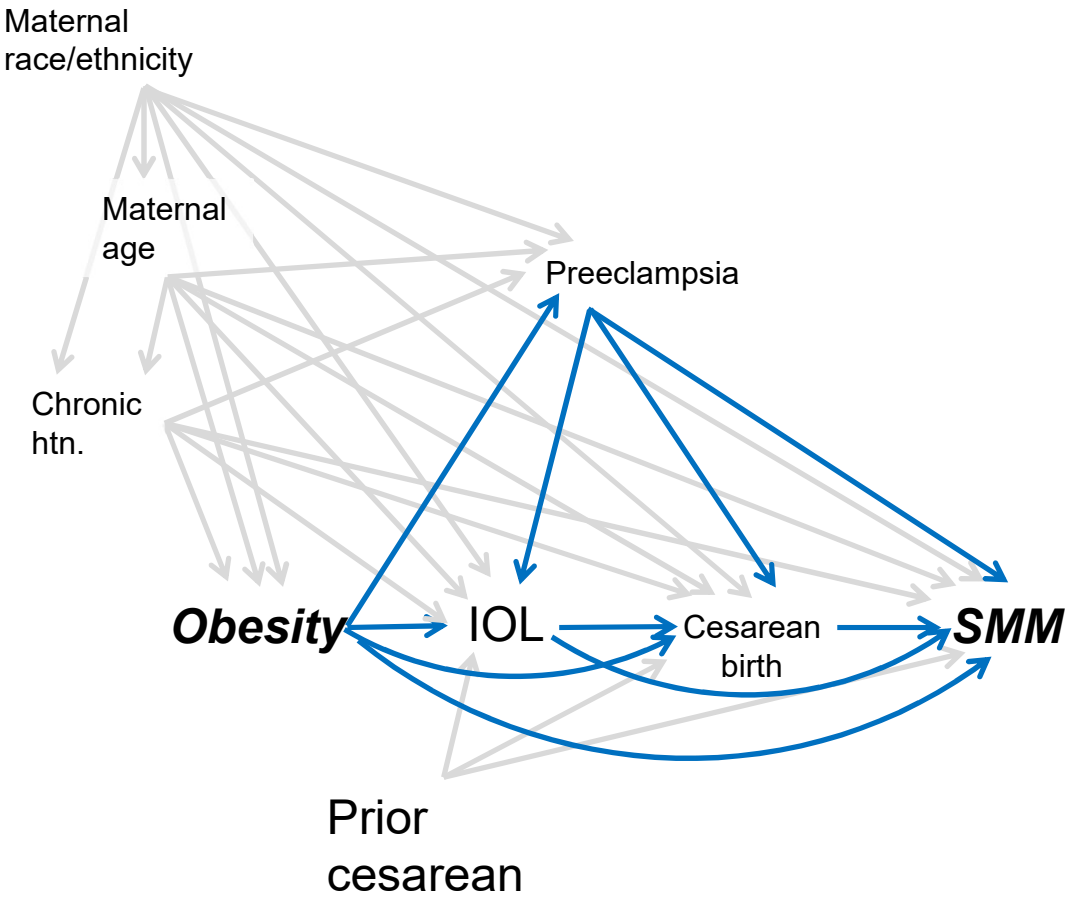


One DAG, many questions



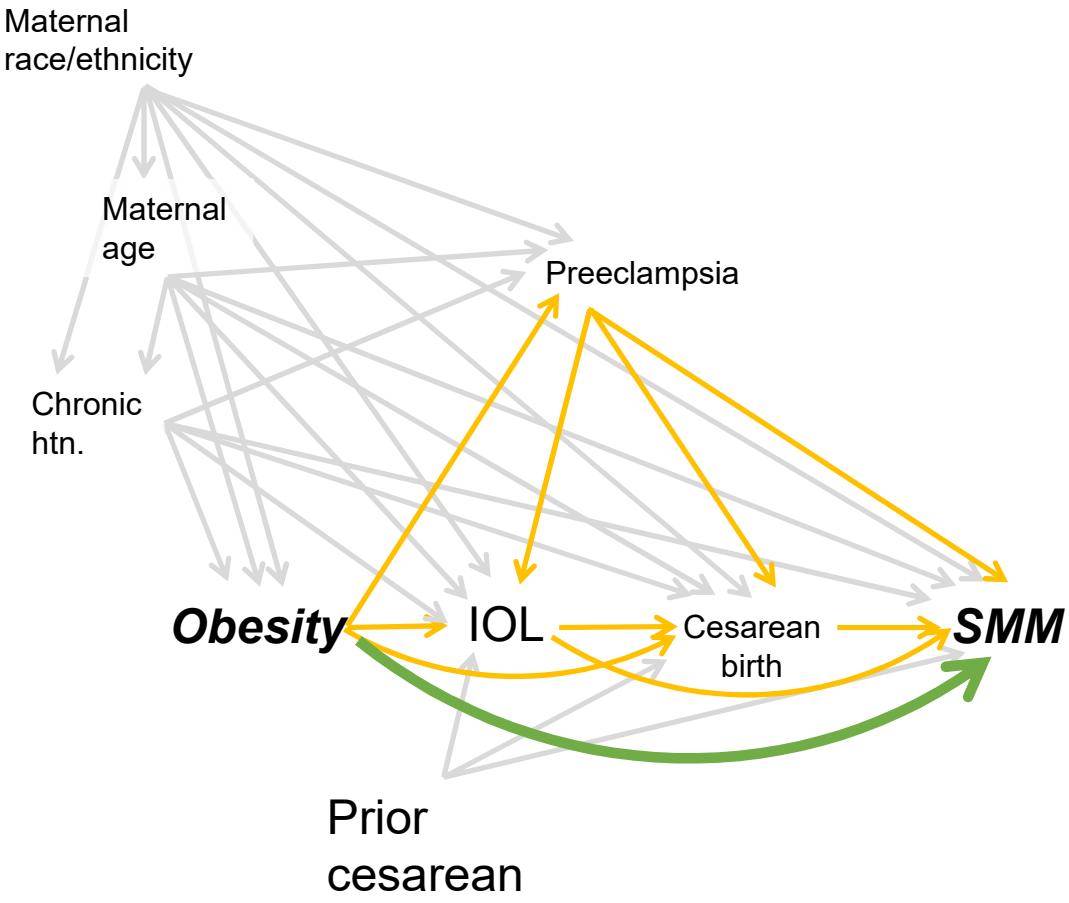
Effects of obstetric procedure use on SMM?

One DAG, many questions



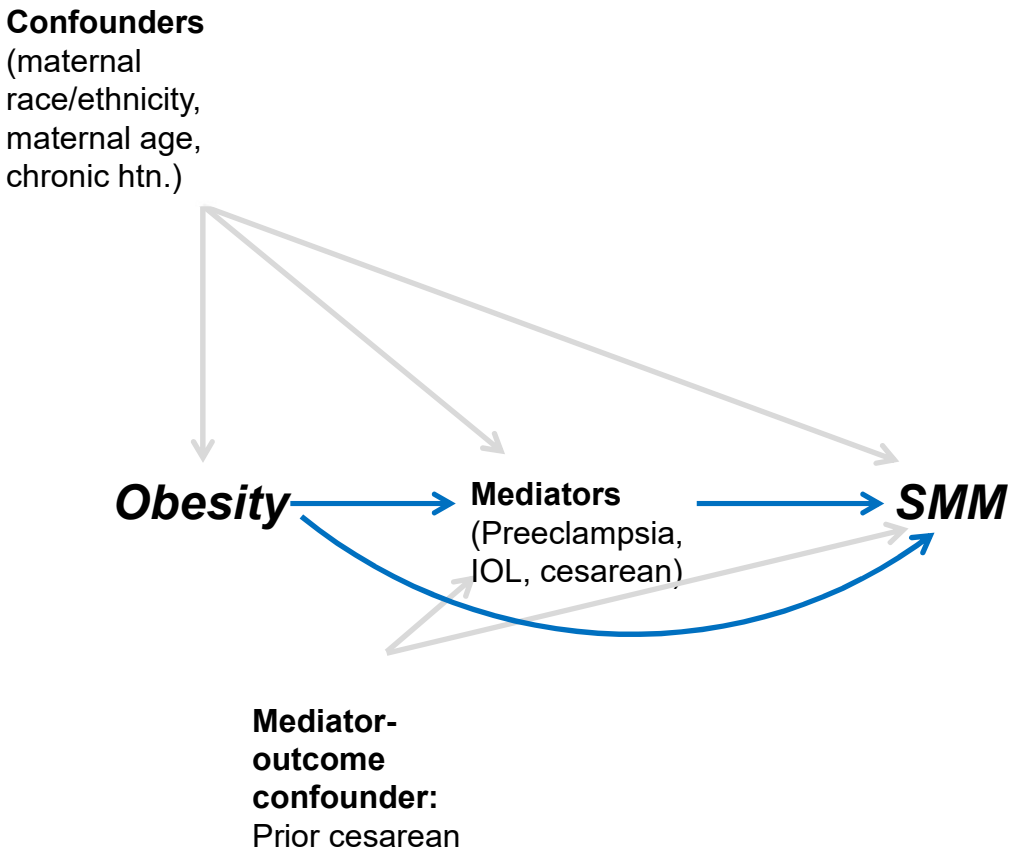
Direct effects *of obesity on SMM?*
Indirect effects *(eg, mediated through morbidities, cesarean)?*

One DAG, many questions



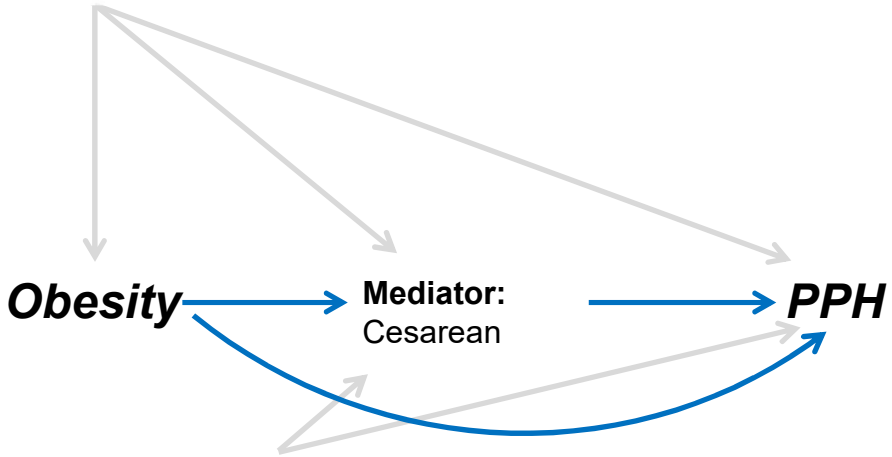
Direct effects *of obesity on SMM?*
Indirect effects *(eg, mediated through morbidities, cesarean)?*

Simplifying the DAG



Simple DAG: simulated data

Confounders:
- Maternal race/ethn.
- Parity
- Maternal age
- Maternal education

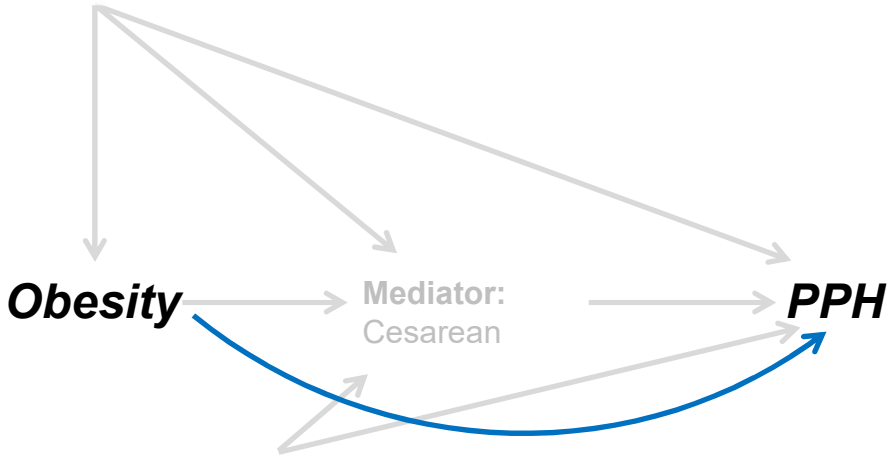


Mediator-outcome confounder:
Prior cesarean

Activity: Familiarize yourself with the data (1)

$$Logit(PPH|Obesity, \bar{W}) = \beta_0 + \beta_1 * Obesity + \beta_2 * nullip. + \cdots + \varepsilon$$

Confounders:
- Maternal race/ethn.
- Parity
- Maternal age
- Maternal education



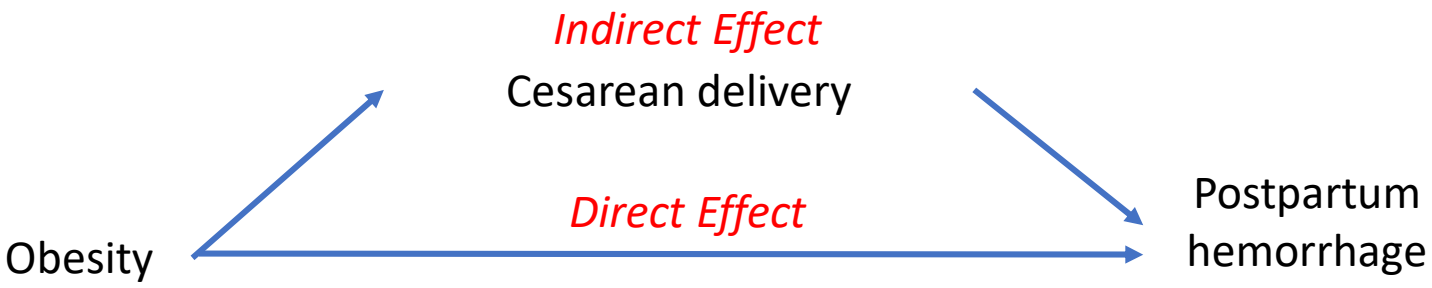
Activity: Familiarize yourself with the data (2)

- Familiarizes yourself with the dataset
  - Google Drive: og_data
- Run some cross-tabs
- Run a basic logistic regression model to determine the total effect of BMI on PPH, controlling for confounders
- Data Dictionary.doc in Google Drive

Agenda

- Question/methods decision tree
- Formulating a question:
  - The effects of obesity on PPH.
- Formal causal mediation analysis
  - Common methods for causal mediation
  - Contemporary thought for causal mediation

Motivating Example:
Obesity → postpartum hemorrhage



Total Effect = Direct Effect + Indirect Effect

Agenda

- Question/methods decision tree
- Formulating a question:
  - The effects of obesity on PPH.
- Formal causal mediation analysis
  - Common methods for causal mediation
  - Contemporary thought for causal mediation

Common Methods for Causal Mediation

- Difference Method:
  - $E[Y|A = a, \bar{W} = \bar{w}] = \beta_0 + \beta_1 A + \beta_2 \bar{W}$
    - Unadjusted model
  - $E[Y|A = a, M = m, \bar{W} = \bar{w}] = \theta_0 + \theta_1 A + \theta_2 M + \theta_3 \bar{W}$
    - Adjusted model
  - Total effect =β_1
  - Direct effect =θ_1
  - Indirect effect =$\beta_1 - \theta_1$
    - Similar to 'proportion explained' $(RR_u - RR_a/RR_u) \cdot 100$
  - No distributions placed on the mediator
    - Binary/discrete/continuous

Common Methods for Causal Mediation

- Product Method (Alwin & Hauser 1975; Baron & Kenny, 1986)
  - $E[M|A = a, \bar{W} = \bar{w}] = \varphi_0 + \varphi_1 A + \varphi_2 \bar{W}$
    - Mediator model
  - $E[Y|A = a, M = m, \bar{W} = \bar{w}] = \theta_0 + \theta_1 A + \theta_2 M + \theta_3 \bar{W}$
    - Adjusted model
  - Direct effect = θ_1
  - Indirect effect = $\varphi_1 \cdot \theta_2$
  - Total effect = $\theta_1 + \varphi_1 \cdot \theta_2$
- Product method and difference method will coincide for continuous outcomes but not binary outcomes (MacKinnon & Dwyer 1993; MacKinnon et al 1995)

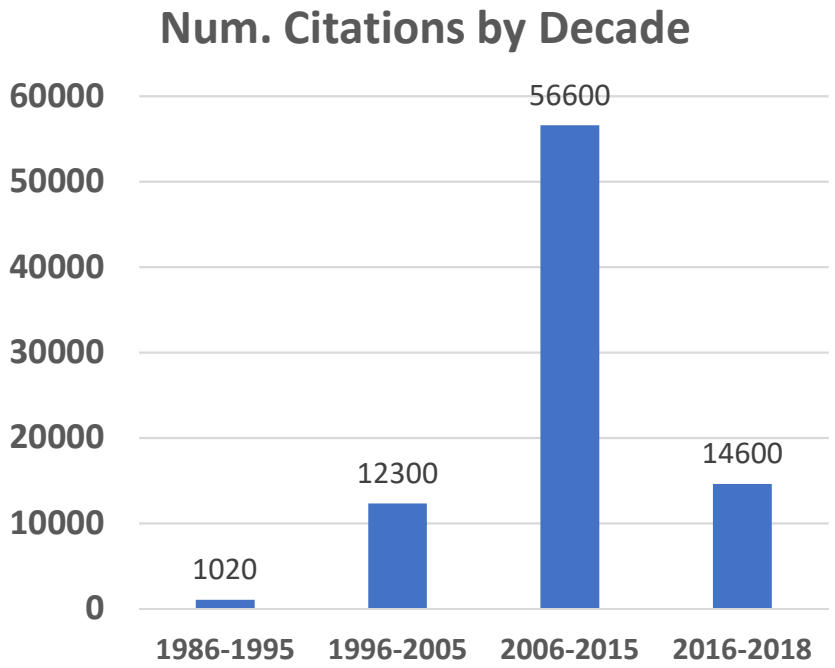
Baron & Kenny (1986)

J Pers Soc Psychol. 1986 Dec;51(6):1173-82.

The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations.

Baron RM, Kenny DA.

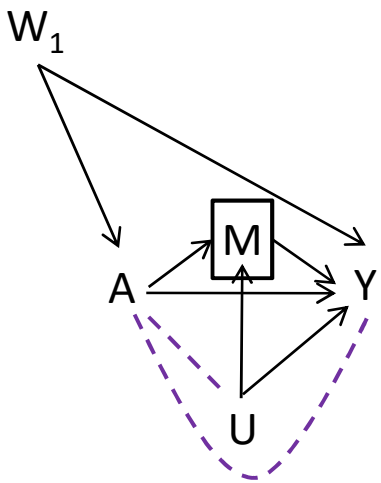
- Cited 84,520 times (Google Scholar 1986 - 2018)
  - ~ 2,561 citations a year for 33 years
  - ~ 7 citations a day
- Adjustment for the mediator
  - "Product method"
- Strong assumptions are required to obtain the direct/indirect effects
- http://davidakenny.net/cm/mediate.htm#CI



Assumptions for Baron & Kenny (1986)

Assumptions/Limitations:

- Linear regression models
  - Estimators for direct and indirect effects are not defined when mediator is binary
  - Recent work has extended this approach to more complex designs
- No exposure-mediator interaction
- No mediator-outcome confounding
  - In Baron & Kenny this is referred to as "Omitted Variable"
  - Not limited to Baron & Kenny – an assumption for any mediation analysis including 'difference method'



Agenda

- Question/methods decision tree
- Formulating a question:
  - The effects of obesity on PPH.
- Formal causal mediation analysis
  - Common methods for causal mediation
  - Contemporary thought for causal mediation

What is FORMAL MEDIATION ANALYSIS?

- Terminology:
  - A collection of tools and processes for identifying, formalizing and quantifying mechanisms (pathway-specific hypotheses)
- Rooted in **potential-outcomes framework**
- Move from association to mechanism

When is a FORMAL MEDIATION ANALYSIS desirable?

- Understand etiology
- Inform intervention strategies
  - Example: Blast → Self-reported hearing difficulties
    - Should Audiology or Mental Health (or both) be involved?
  - Example: BMI → Severe maternal morbidity
    - Support vaginal birth among women with high BMI[1]
- In absence of total effect, mediated effect may be informative
  - Direct and mediated effects have opposite signs[2]
- Strengthen the evidence the total effect is causal

[1]*Lenoard SA, Carmichael SL, Main EK, Lyell DJ, Abrams B (2019). Risk of severe maternal morbidity in relation to prepregnancy body mass index: Roles of maternal co-morbidities and caesarean birth, Paediatr Perinat Epidemiol, 00: 1-9.*
[2]*MacKinnon, D. P. (2008). Introduction to statistical mediation analysis. New York, NY: Erlbaum.*

Salient Assumptions in Causal Mediation Analysis

- Moving from observational research to potential outcomes for the purposes of causal inference:
  - We observe this: $P(Y = 1|A = 1) - P(Y = 1|A = 0)$
  - But what we want is this: $P(Y_{a=1} = 1) - P(Y_{a=0} = 1)$

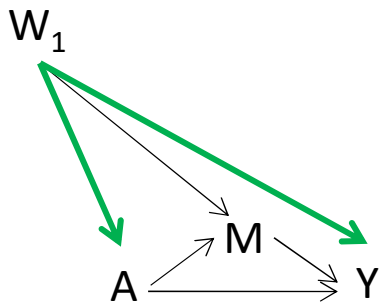
- We relate the two above through <u>assumptions</u>:

1. Exchangeability
   - Mediator-outcome confounder not affected by exposure
2. Positivity
3. Consistency

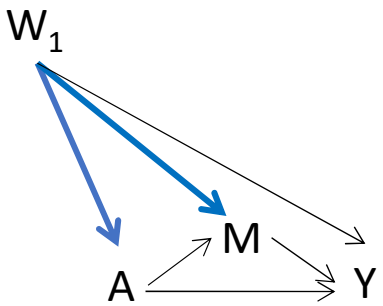
Salient Assumptions in Causal Mediation Analysis: (1) *Exchangeability*

- Exchangeability (Sequential Ignorability)
  - No residual or unmeasured confounding, including **mediator – outcome** confounding.
  - While randomization of the treatment in RCT minimizes A→Y confounding, M→Y associations to not similarly benefit

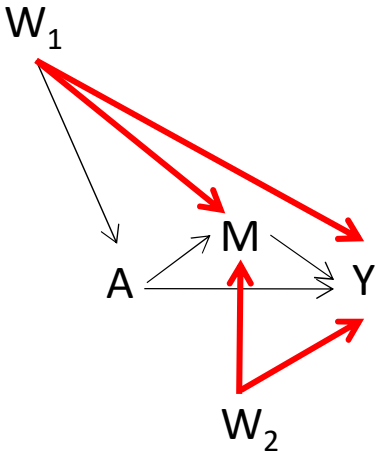
A → Y



A → M

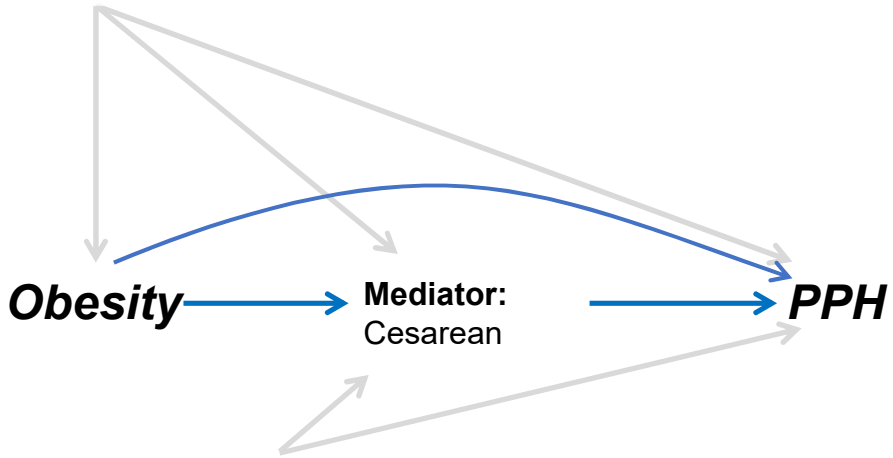


M → Y



Salient Assumptions in Causal Mediation Analysis: (1) *Exchangeability*

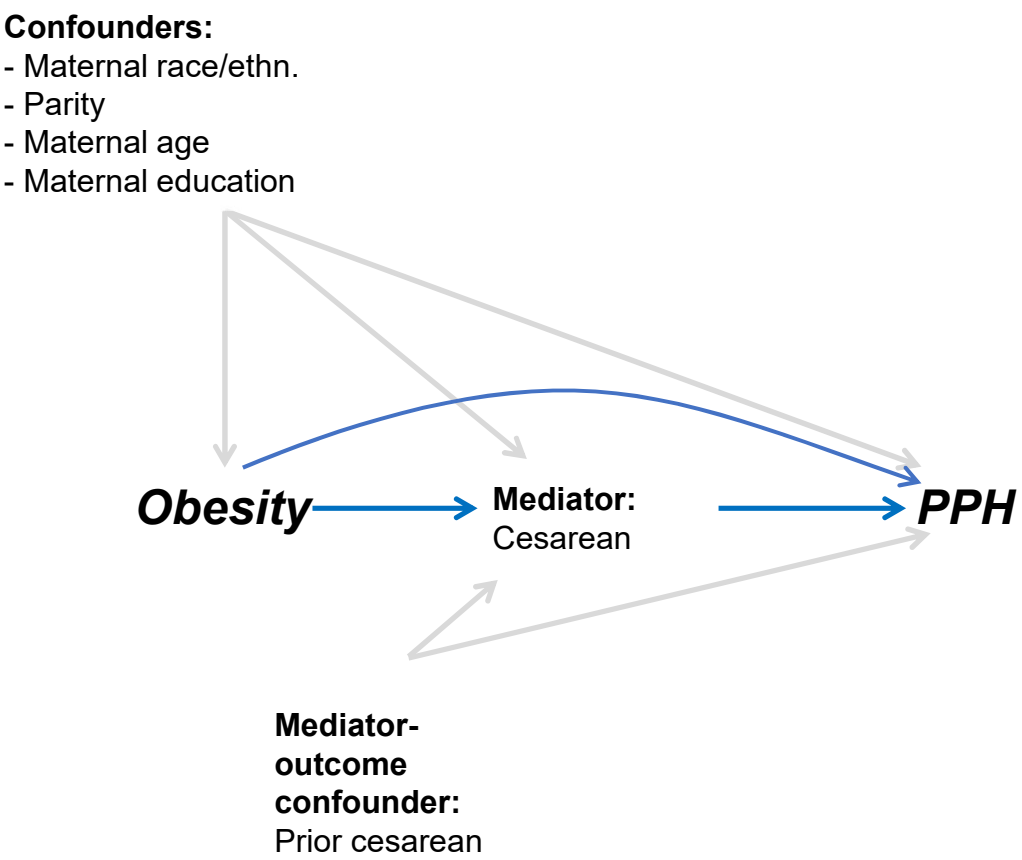
Confounders:
- Maternal race/ethn.
- Parity
- Maternal age
- Maternal education



Mediator-outcome confounder:
Prior cesarean

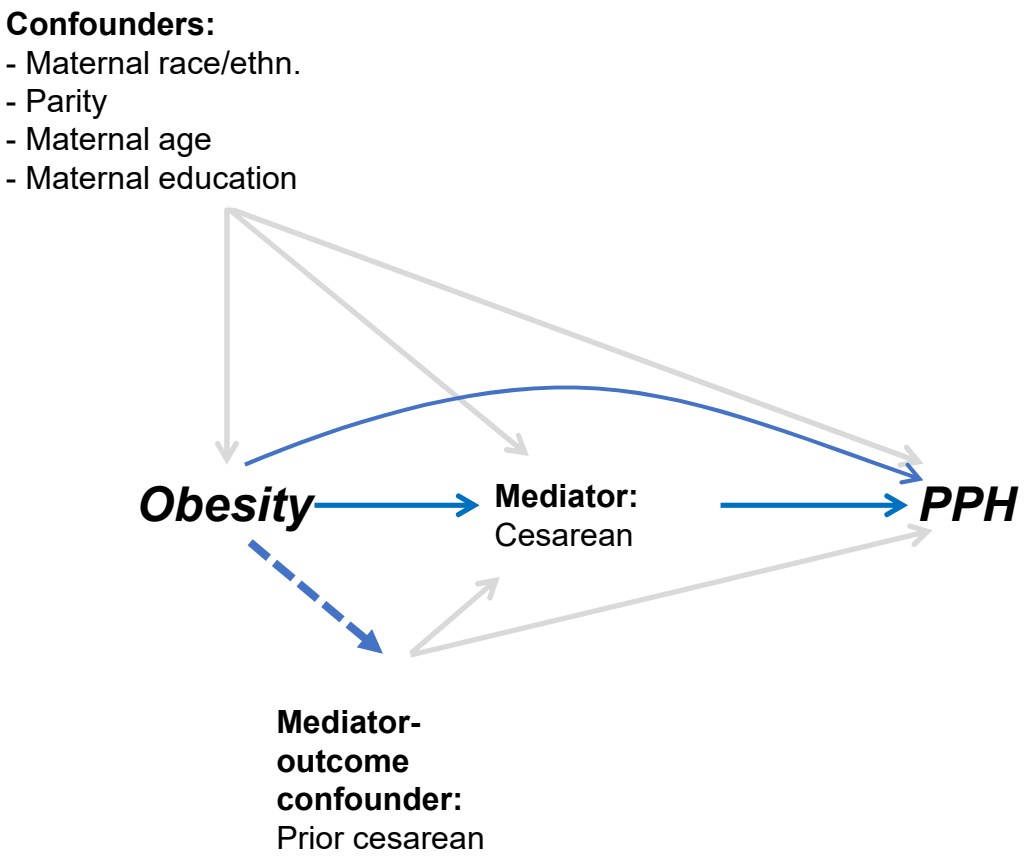
Salient Assumptions in Causal Mediation Analysis: (1) *Exchangeability*

- Mediator-outcome confounder not affected by exposure



Salient Assumptions in Causal Mediation Analysis: (1) *Exchangeability*

- If they are intertwined, we have confounding by a causal intermediate
- Similar structure to time-dependent confounding

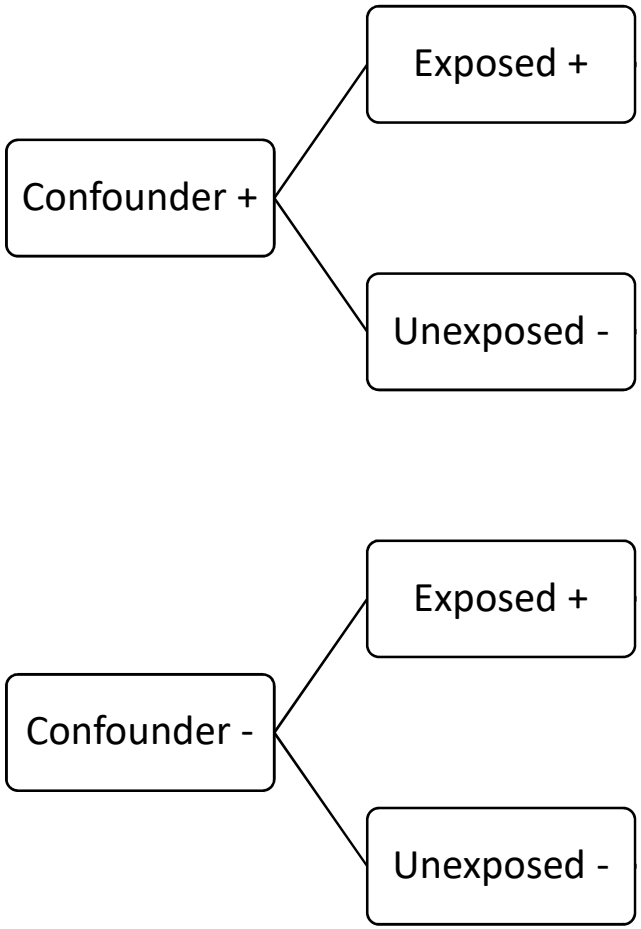


Salient Assumptions in Causal Mediation Analysis: (2) *Positivity*

- For any values of confounders, all exposure values must have a non-zero probability.[1]
- For any value of confounders and exposure, all **mediator values** must have a non-zero probability.[1]
- A testable assumption through 2x2 tables

[1] Lange, Hansen, Sorensen & Galatius, 2017

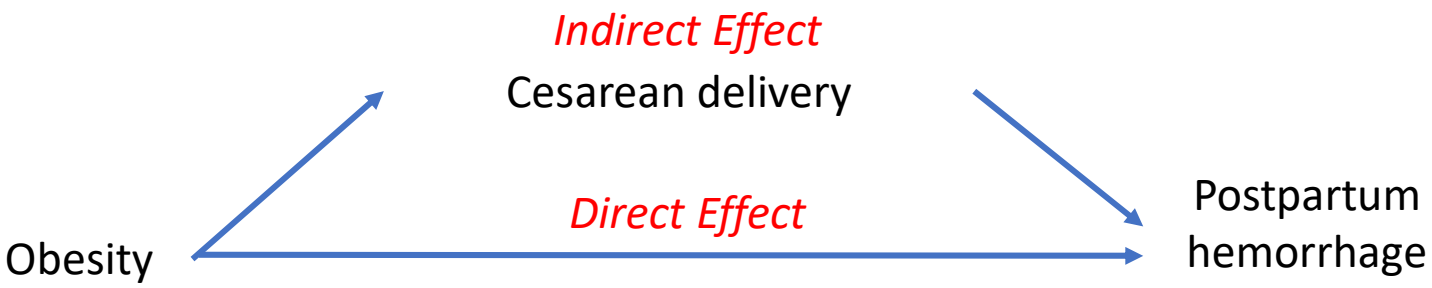
Salient Assumptions in Causal Mediation Analysis: (2) *Positivity*



Salient Assumptions in Causal Mediation Analysis: (3) *Consistency*

- One needs to be able to explain how a certain level of exposure **and mediator** could be hypothetically assigned to a person exposed to a different level.
  - The unobserved, new exposure and mediator causally acting on outcome, producing an unobserved, counterfactual outcome
  - Concordance of the counterfactual outcome with something we would observe in the real world.

Motivating Example:
Obesity → postpartum hemorrhage



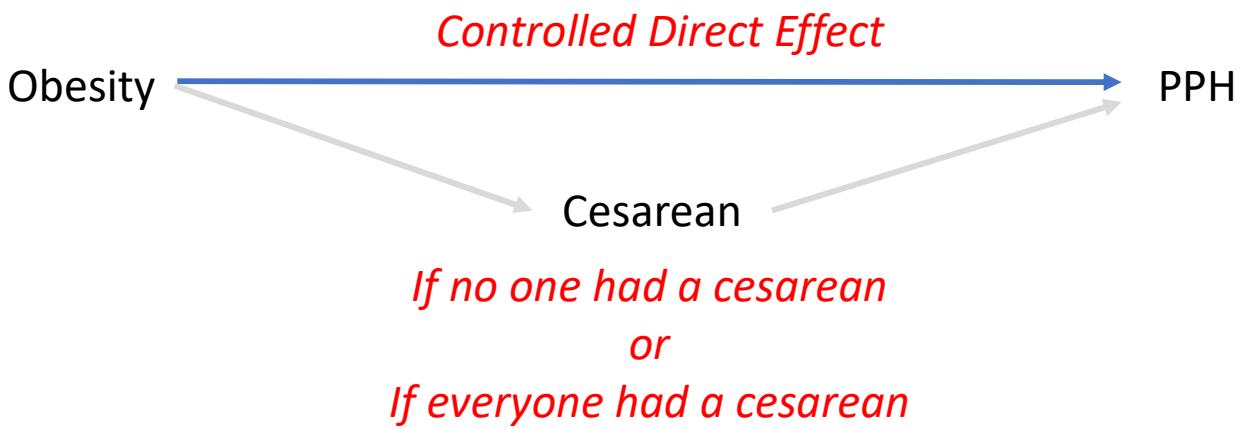
Total Effect = Direct Effect + Indirect Effect

Formal Causal Mediation Terminology: *Direct/Indirect Effects*

- Direct Effect:
  - not mediated by an intermediate variable
  - A → Y
- Indirect Effect:
  - effects are relayed through an intermediate variable
  - A → M → Y
- Total Effect:
  - Direct Effect + Indirect Effect

Formal Causal Mediation Terminology: *Controlled Direct Effects*

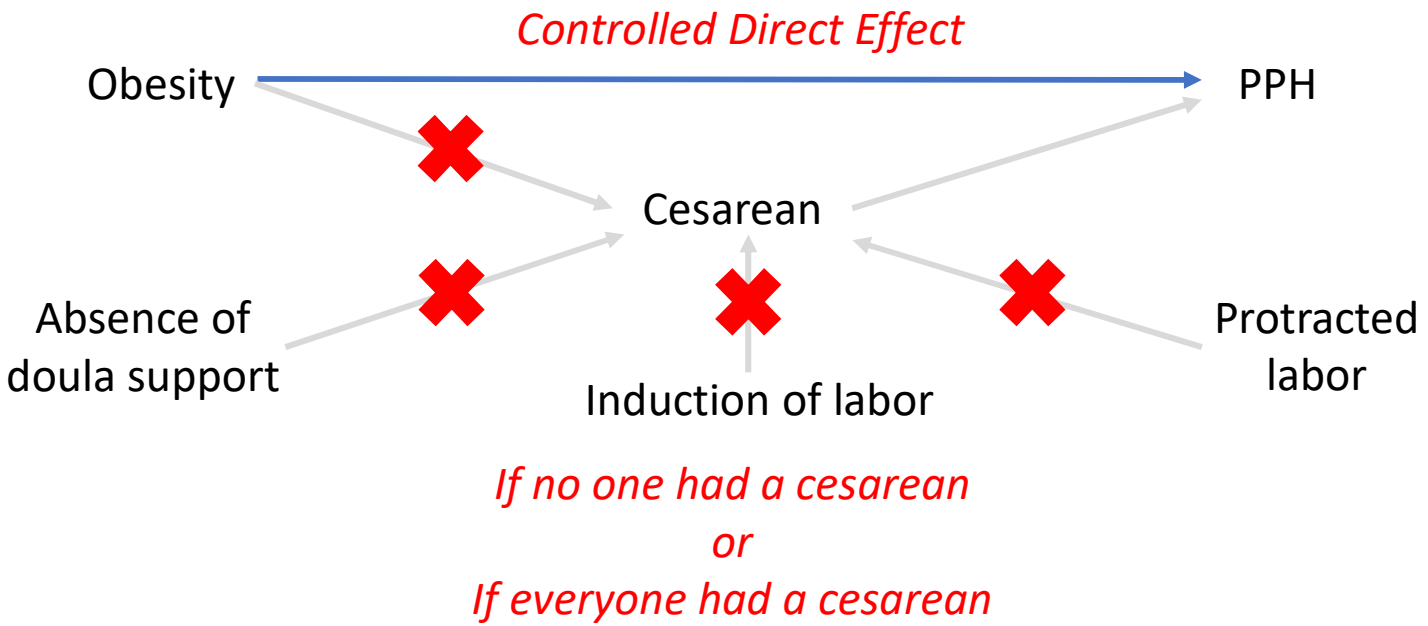
- Effect of exposure on outcome that would be observed if the mediator were controlled or set to a fixed value. [1]



[1]Petersen ML, Sinisi SE, van der Laan MJ (2006). *Estimation of direct causal effects,* Epidemiology, 17 (3), 276-284

Formal Causal Mediation Terminology: *Controlled Direct Effects*

- Effect of exposure on outcome that would be observed if the mediator were controlled or set to a fixed value. [1]



[1]Petersen ML, Sinisi SE, van der Laan MJ (2006). *Estimation of direct causal effects,* Epidemiology, 17 (3), 276-284

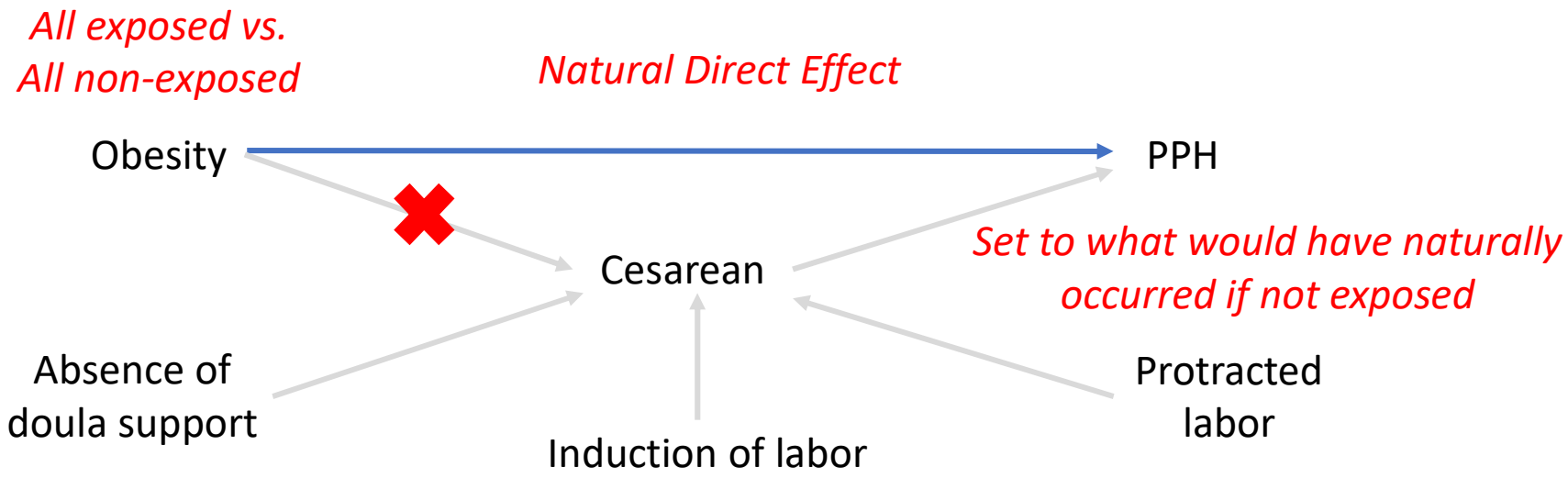
Formal Causal Mediation Terminology

- May not seem realistic to think of the mediator being the same for all subjects – may be more realistic to let the mediator **naturally** vary
- The mediator takes on the value it would have **naturally** if the exposure had not occurred.[1]

[1]Petersen ML, Sinisi SE, van der Laan MJ (2006). *Estimation of direct causal effects,* Epidemiology, 17 (3), 276-284

Formal Causal Mediation Terminology: *Natural Direct Effects*

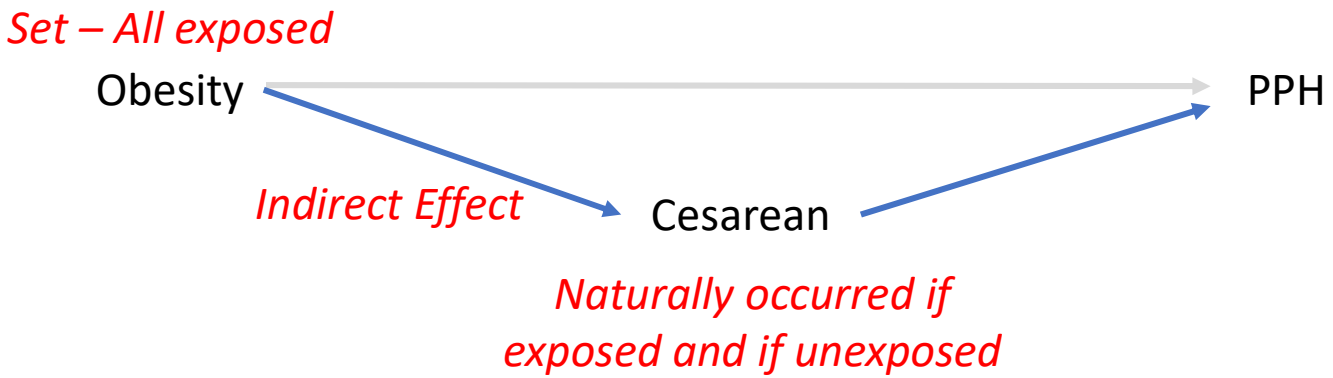
- Effect of exposure on outcome that would be observed if the mediator were set to the value it would have naturally have taken in the absence of the exposure.[1]



[1]Petersen ML, Sinisi SE, van der Laan MJ (2006). *Estimation of direct causal effects,* Epidemiology, 17 (3), 276-284

Formal Causal Mediation Terminology: *Natural Indirect Effects*

- If the mediator were somehow changed to what it would be without the exposure
  - The exposure is set
- Estimates how much the outcome would change if the exposure acted only through modifying the mediator.[1]



Formal Causal Mediation Terminology

- Natural direct/indirect effects do not presume no interactions between exposure and mediator on outcome
- Variation in the mediator level enables effect decomposition of the total effect into a natural direct and indirect effects.
- No analogous definition of “controlled indirect effect”
  - Generally, controlled direct effects are not useful for effect decomposition

Counterfactual Framework

Remember, for binary exposure indicator A, the familiar Y of associational regression analysis is replaced with:

- Y_1 (units potential outcome when exposed or A = 1)
- Y_0 (units potential outcome when unexposed or A = 0)
- Y_a when setting A = a

- Same concepts for Mediator:
  - M_1 (units potential mediator when exposed or A = 1)
  - M_0 (units potential mediator when unexposed or A = 0)
  - M_a when setting A = a
- Combined:
  - $Y_{a,m}$ units potential outcome when setting A = a and M = m
- Nested:
  - $Y_{a,m(a)}$ units potential outcome when setting A = a and M takes on value had A = a

Counterfactual Framework

- ***Controlled Direct Effects (CDE):***
  - $CDE = Y_{1,m}$ - $Y_{0,m}$
  - where M is fixed at m (CDE depends on level of m)
- ***Natural Direct Effects (NDE):***
  - NDE = $Y_{1,m(0)}$ - $Y_{0,m(0)}$
  - where M is set at M(0)
- ***Natural Indirect Effects (NIE):***
  - NIE = $Y_{1,m(1)}$ - $Y_{1,m(0)}$
- ***Total (Average) Causal Effect:*** *NDE + NIE*

Counterfactual Framework
Measures of Association: Odds Ratios[1]

Controlled Direct Effect: Comparing A=1 to A=0 setting M=m

$$OR = \frac{P\left(Y_{1,m} = 1\right)/P\left(Y_{1,m} = 0\right)}{P\left(Y_{0,m} = 1\right)/P\left(Y_{0,m} = 0\right)}$$

[1] VanderWeele TJ and Vansteelandt S (2010). *Odds ratios for mediation analysis for a dichotomous outcome.* AJE, 172(12): 1339 - 1348

Counterfactual Framework
Measures of Association: Odds Ratios[1]

Natural Direct Effect: *Comparing A=1 to A=0 setting M=M_0*

$$OR = \frac{P\left(Y_{1,m(0)} = 1\right)/P\left(Y_{1,m(0)} = 0\right)}{P\left(Y_{0,m(0)} = 1\right)/P\left(Y_{0,m(0)} = 0\right)}$$

Natural Indirect Effect: *Comparing M=M_1 to M=M_0 setting A=1*

$$OR = \frac{P\left(Y_{1,m(1)} = 1\right)/P\left(Y_{1,m(1)} = 0\right)}{P\left(Y_{1,m(0)} = 1\right)/P\left(Y_{1,m(0)} = 0\right)}$$

Total Causal Effect

$$OR = OR^{NDE} \cdot OR^{NIE}$$

[1] VanderWeele TJ and Vansteelandt S (2010). *Odds ratios for mediation analysis for a dichotomous outcome.* AJE, 172(12): 1339 - 1348

Notation

- Exchangeability (Conditional)
  - $Y_{a,M_a} \perp A|\bar{W}$
  - $M_a \perp A|\bar{W}$
  - $Y_{a,m} \perp M|(A,\bar{W})$
- Positivity
  - For any value of confounder, all exposure values have a non-zero probability AND for any value of confounder/exposure, all mediators have a non-zero probability
  - $P\ (A = a|\bar{W} = \bar{w}) > 0$ for all a, w
  - $P(M = m|A = a, \bar{W} = \bar{w}\ ) > 0$ for all a, w, m

Notation

- Consistency
  - Nested counterfactual takes the observed value when the treatment and mediator are set to the value they would have naturally have had in the absence of intervention
  - $P\left(Y_{a,m} = Y\right) = 1$ and $P(M_a = M) = 1$

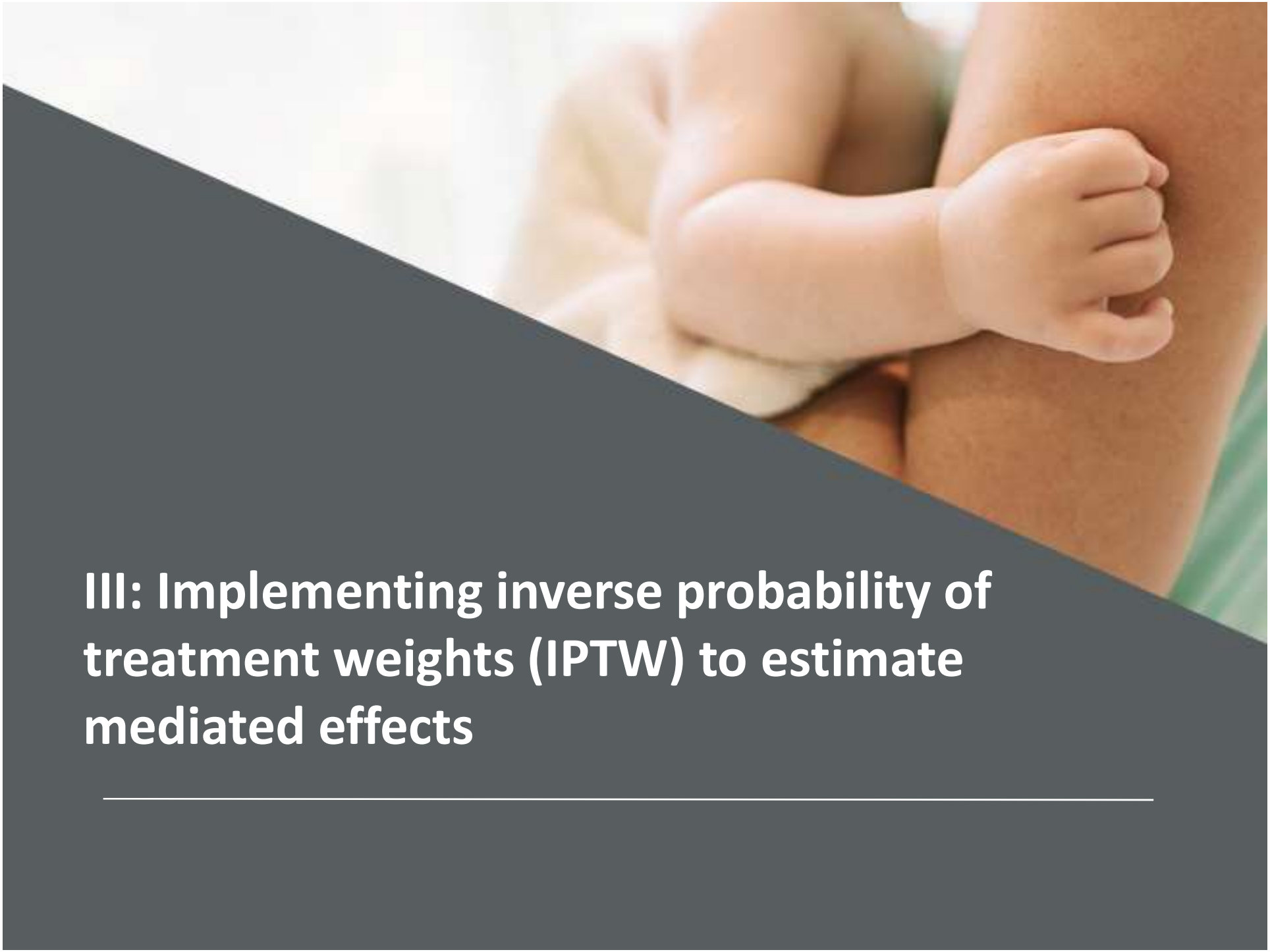
Computation/Implementation

- SAS
  - proc causalmed
  - https://support.sas.com/documentation/onlinedoc/stat/143/causalmed.pdf
  - https://video.sas.com/detail/video/5802737116001/introducing-the-causalmed-procedure-for-causal-mediation-analysis
- STATA
  - paramed
  - idecomp
  - medeff (medsens)
  - gformula
  - https://www.stata.com/meeting/italy13/abstracts/materials/it13_grotta.pdf
- R
  - mediation
  - https://cran.r-project.org/web/packages/mediation/vignettes/mediation.pdf
  - medflex

Break

When we return:

Implementing inverse probability of treatment weights (IPTW) to estimate mediated effects



III: Implementing inverse probability of treatment weights (IPTW) to estimate mediated effects

Agenda

- Recall motivation for using IPTW (decision tree)
- IPTW theory and mechanics
- Other applications of inverse probability weights (IPW):
  - Censoring/selection bias (IPCW)
  - Time-dependent confounding (IPTW)
- Data analysis activity

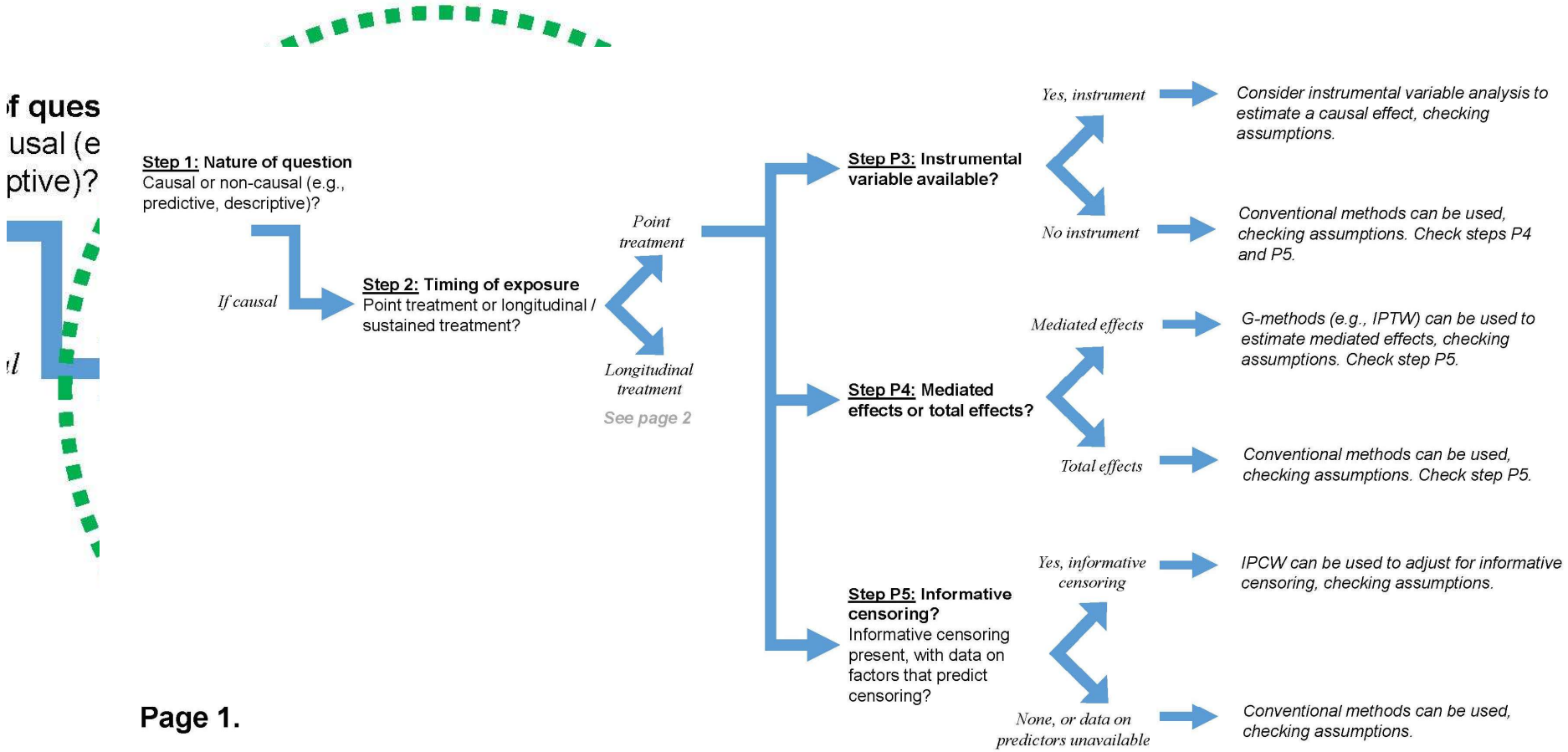
Agenda

- Recall motivation for using IPTW (decision tree)
- IPTW theory and mechanics
- Other applications of inverse probability weights (IPW):
  - Censoring/selection bias (IPCW)
  - Time-dependent confounding (IPTW)
- Data analysis activity

Recall the causal question

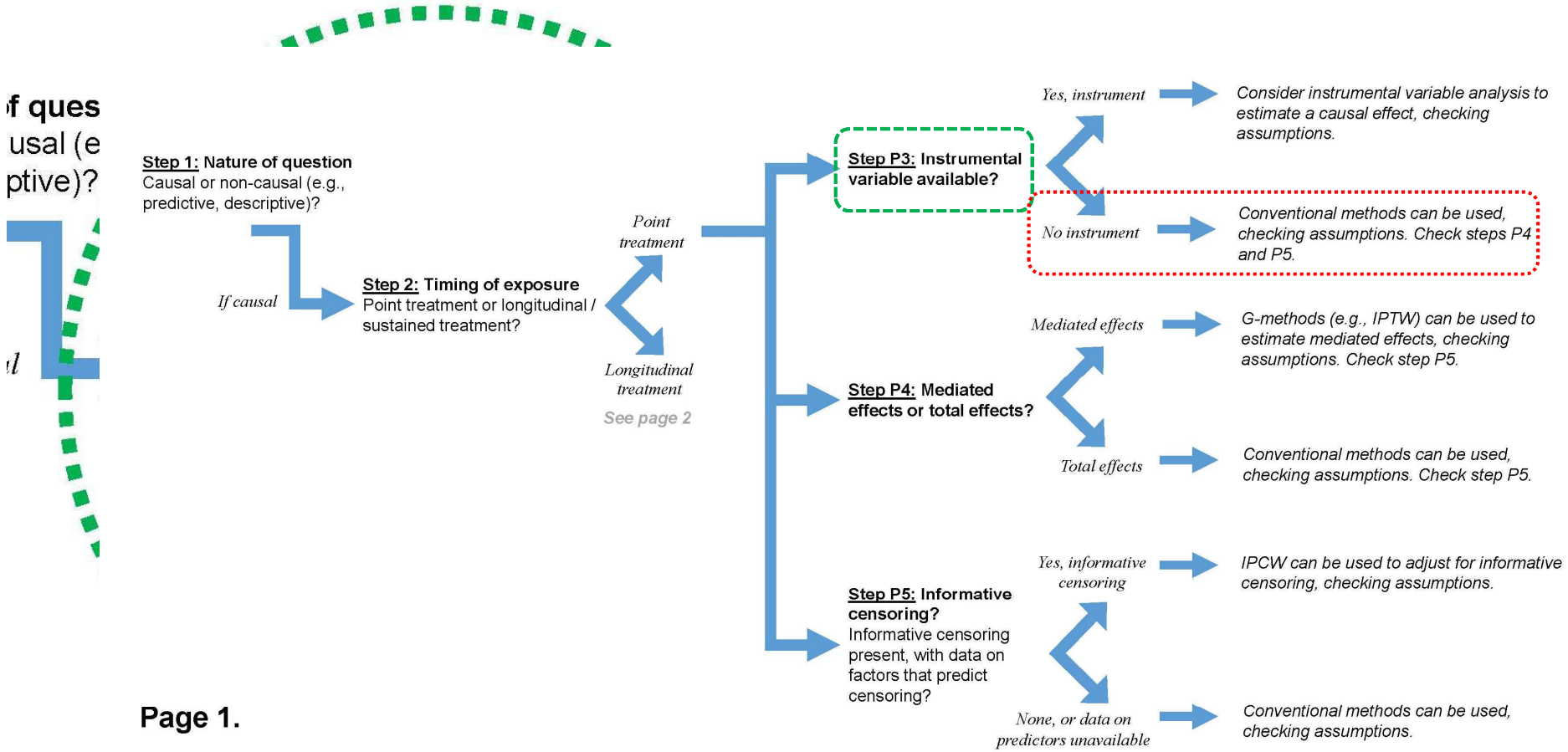
- What is the direct effect of obesity on PPH, not mediated through cesarean birth?
  - What proportion is mediated through cesarean birth?

Question/methods decision tree

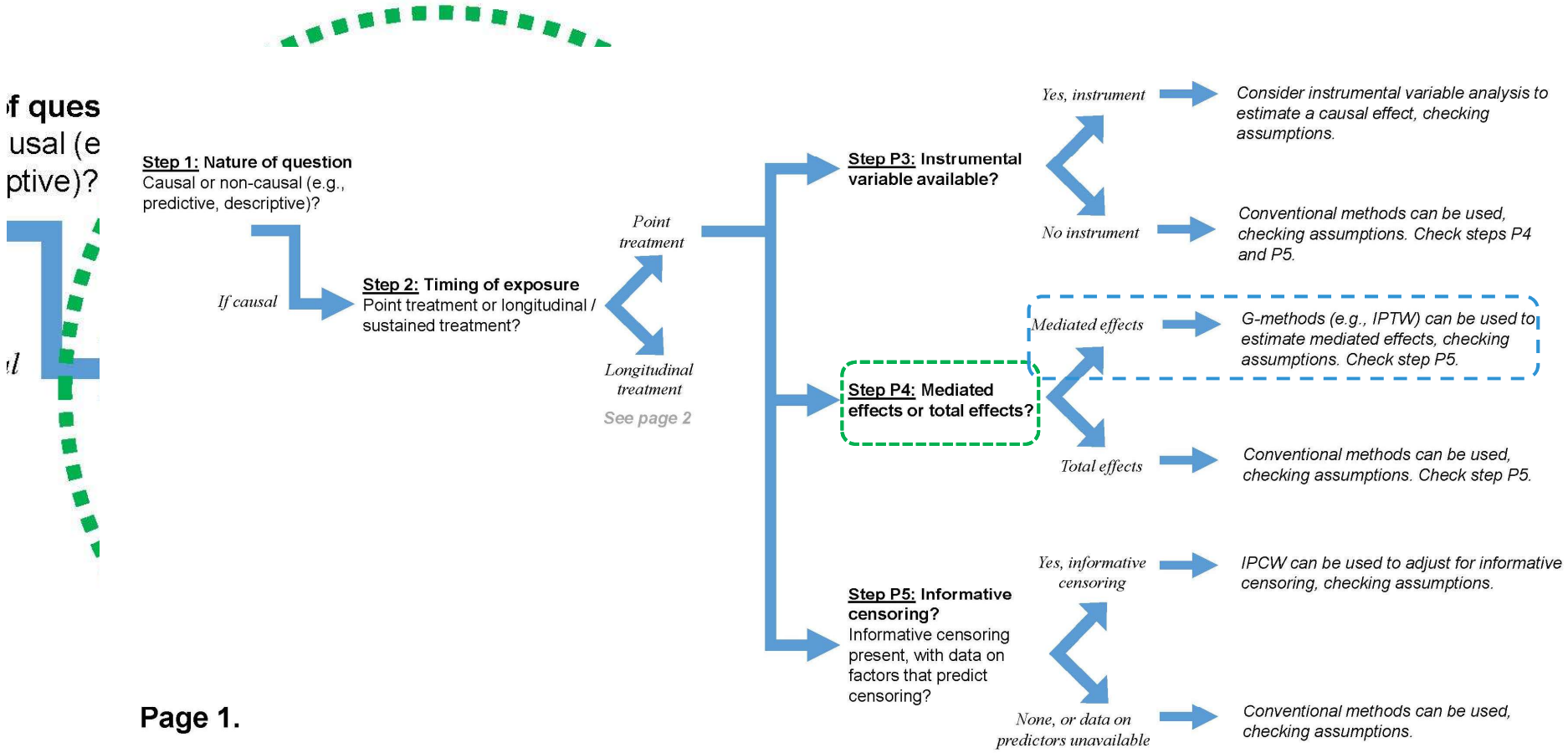


Page 1.

Question/methods decision tree



Decision tree: **Mediated effects**



Page 1.

Agenda

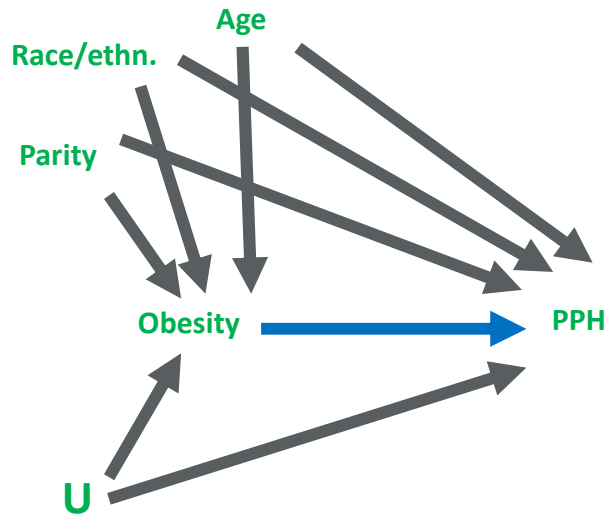
- Recall motivation for using IPTW (decision tree)
- IPTW theory and mechanics
- Other applications of inverse probability weights (IPW):
  - Censoring/selection bias (IPCW)
  - Time-dependent confounding (IPTW)
- Data analysis activity

The missing data problem of CI

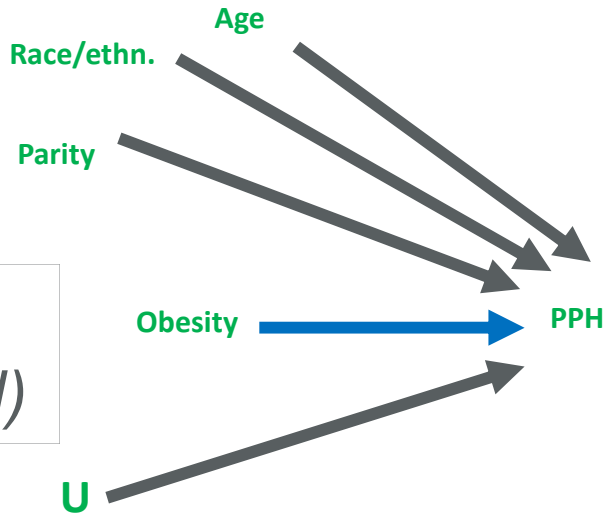
- *We can only observe each study unit under one exposure setting.*
- *Therefore, only one of the potential outcomes may be observed for each study unit.*

Confounding as a missing data problem

- For any given research question and dataset, assume the existence of a "full data set."
- A hypothetical ideal: a dataset in which there is no confounding, and exposure is unassociated with any other variables.
- For obesity and PPH, imagine a dataset where obesity is independent from:
  - Race/ethnicity
  - Parity
  - Age

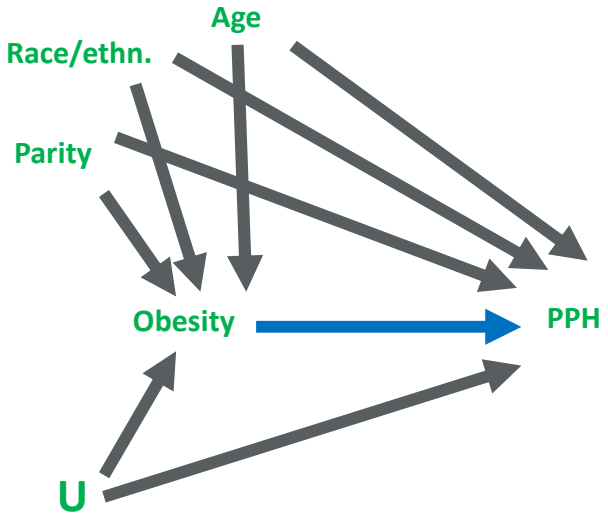


Confounding as a missing data problem



Full data (hypothetical)

Covariates not associated with exposure, thus no confounding



Observed data (the real world)

Confounding as a missing data problem

- Assume that the observed data *are a subset* of the full data (i.e., a sample).
- Our observed data are a subset in which some individuals are over-represented and under-represented, relative to their share of the full population.

Observed data

	Obesity	No obesity
Latina (%)	0.6	0.5
Nullip. (%)	0.35	0.45
Age (y)	31	27
…		

Full data

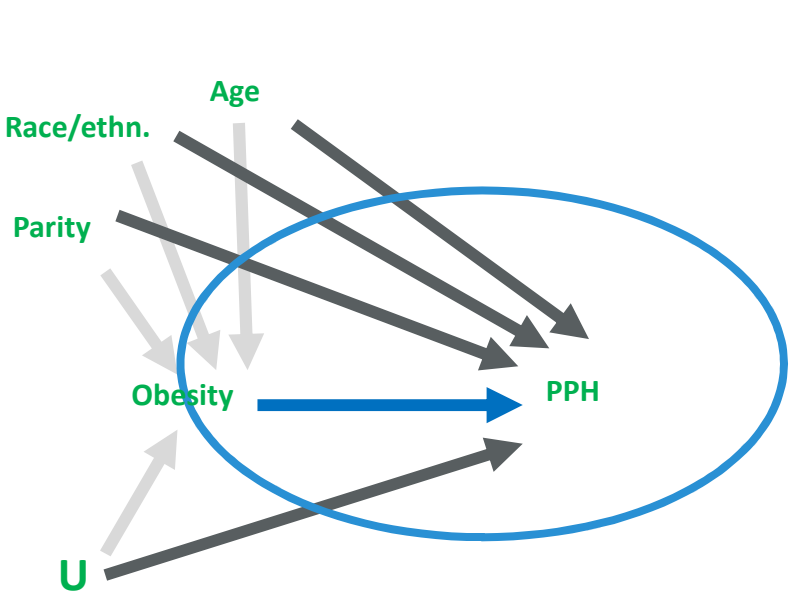
	Obesity	No obesity
Latina (%)	0.55	0.55
Nullip. (%)	0.4	0.4
Age (y)	29	29
…		

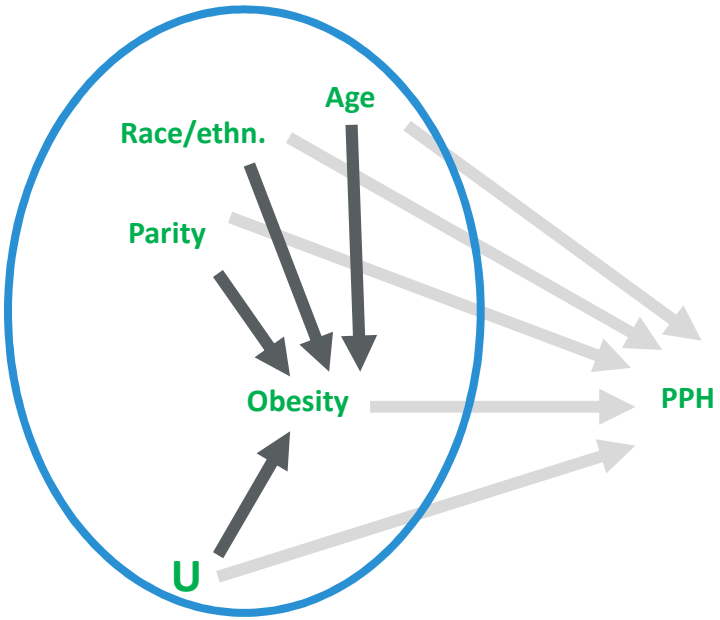
Confounding as a missing data problem

- Assume that the observed data *are a subset* of the full data (i.e., a sample).
- Our observed data are a subset in which some individuals are over-represented and under-represented, relative to their share of the full population.
- Inverse probability weighting is a weighted regression approach that up-weights/down-weights the observed data to approximate the full population.
- Other weighted regression approaches:
  - Sampling weights for complex survey sampling
  - Inverse-variance weighting for meta-analysis
- Terms you may encounter: “ghost population,” pseudo-population, …

The propensity score

- IPW builds on propensity score methods.
- Propensity score defined as the probability of being exposed, given covariates:
  - $\Pr(A = 1|\bar{W} = \bar{w})$
- This is also referred to as the "treatment mechanism": the factors affecting how people received exposure.
- This contrasts to what we normally model: the outcome.





Treatment mechanism

- In RCTs, the treatment mechanism is known, because it was designed by investigators.
  - Often, it is as simple as $\Pr(A = 1) = 0.5$
- In observational research, the treatment mechanism is unknown.
  - We model the treatment mechanism in IPW – this model is called the *treatment model*.
- Accurate identification of the treatment mechanism would enable causal inference.
- This is why causal inference is more straightforward in RCTs.

IPW as a propensity score estimator

- IPW is thought of as being 1 of the 4 PS estimators.
  - Weighting
  - Matching
  - Stratification
  - Multivariable adjustment
- The key difference:
  - PS models the probability of being *treated*.
  - IPW models the probability of received *treatment actually received*.
    - For exposed people, the IPW uses the PS: $\Pr(A = 1|\overline{W} = \overline{w})$
    - For unexposed people, the IPW uses probability of being *untreated*: $\Pr(A = 0|\overline{W} = \overline{w})$ = [1 - $\Pr(A = 1|\overline{W} = \overline{w})$]

Implementing IPW to adjust for confounding: Steps 1-2

1) Fit a treatment model.
   - e.g., $Logit(A|\bar{W}) = \beta_0 + \beta_1 * W^1 + \beta_2 * W^2 + \cdots + \varepsilon$
2) Calculate conditional probability of treatment:

   $\Pr(A = 1|\bar{W} = \bar{w})$

   $\Pr(Obesity = 1|Parity, age, race/ethn)$

   - This probability (0<Pr<1) quantifies how likely obesity was, given covariate profile.

Implementing IPW to adjust for confounding: Steps 3-4

1) Fit a treatment model.
   - e.g., $Logit(A|\bar{W}) = \beta_0 + \beta_1 * W^1 + \beta_2 * W^2 + \cdots + \varepsilon$
2) Calculate conditional probability of treatment.
3) Calculate weights (IPW)
4) Fit outcome model (e.g., MSM).
   - e.g., $Logit(Y_a) = \beta_0' + \beta_1' * a + \varepsilon$; weights=IPW

IPW formula: unstabilized

- IPW (unstabilized):
  - Treated:

$$IPW = \frac{1}{\Pr(A = 1|\bar{W} = \bar{w})}$$

  - Untreated:

$$IPW = \frac{1}{\Pr(A = 0|\bar{W} = \bar{w})}$$

IPW formula

- IPW (unstabilized):
  - Generally:

$$IPW$$

$$= \frac{I(A=1)}{\Pr(A=1|\bar{W}=\bar{w})} + \frac{1-I(A=1)}{\Pr(A=0|\bar{W}=\bar{w})}$$

Stabilized IPW formula

- Using ***stabilized weights*** decreases extreme weight values, increases efficiency (i.e., lower SE).
- The marginal probability of exposure is the numerator, instead of 1:
  - Recall unstabilized IPW (treated):

$$IPW = \frac{1}{\Pr(A = 1|\bar{W} = \bar{w})}$$

  - Stabilized IPW (treated):

$$IPW = \frac{\Pr(A = 1)}{\Pr(A = 1|\bar{W} = \bar{w})}$$

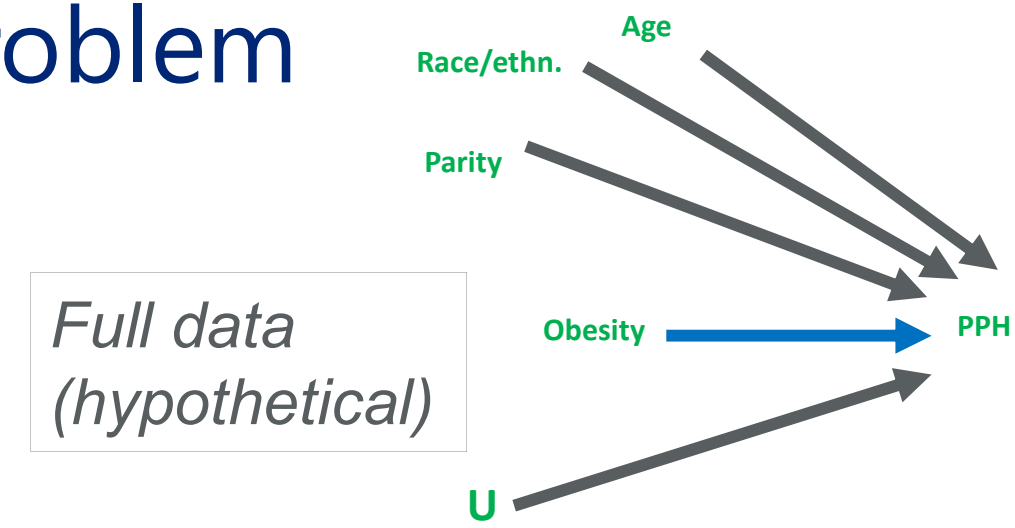
Stabilized IPW formula

- Stabilized IPW, generally:

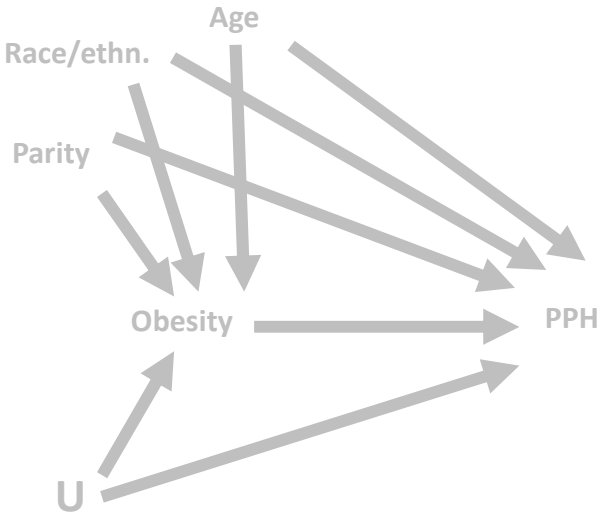
$$IPW$$

$$= \frac{I(A=1) * \Pr(A=1)}{\Pr(A=1|\overline{W}=\overline{w})} + \frac{[1-I(A=1)] * \Pr(A=0)}{\Pr(A=0|\overline{W}=\overline{w})}$$

IPTW resolved the missing data problem



Full data (hypothetical)



Observed data (the real world)

Fit weighted outcome model, using IPWs

- *Marginal structural model:* compares outcome if everyone were exposed, versus everyone unexposed.
- Results are analogous to marginal RR:

$$E(Y_{a=1}) / E(Y_{a=0})$$

- The MSM takes the form:

$$Logit(Y_a) = \beta_0' + \beta_1' * a + \varepsilon; \text{ weights=IPW}$$

  - β' used to denote different, marginal interpretation

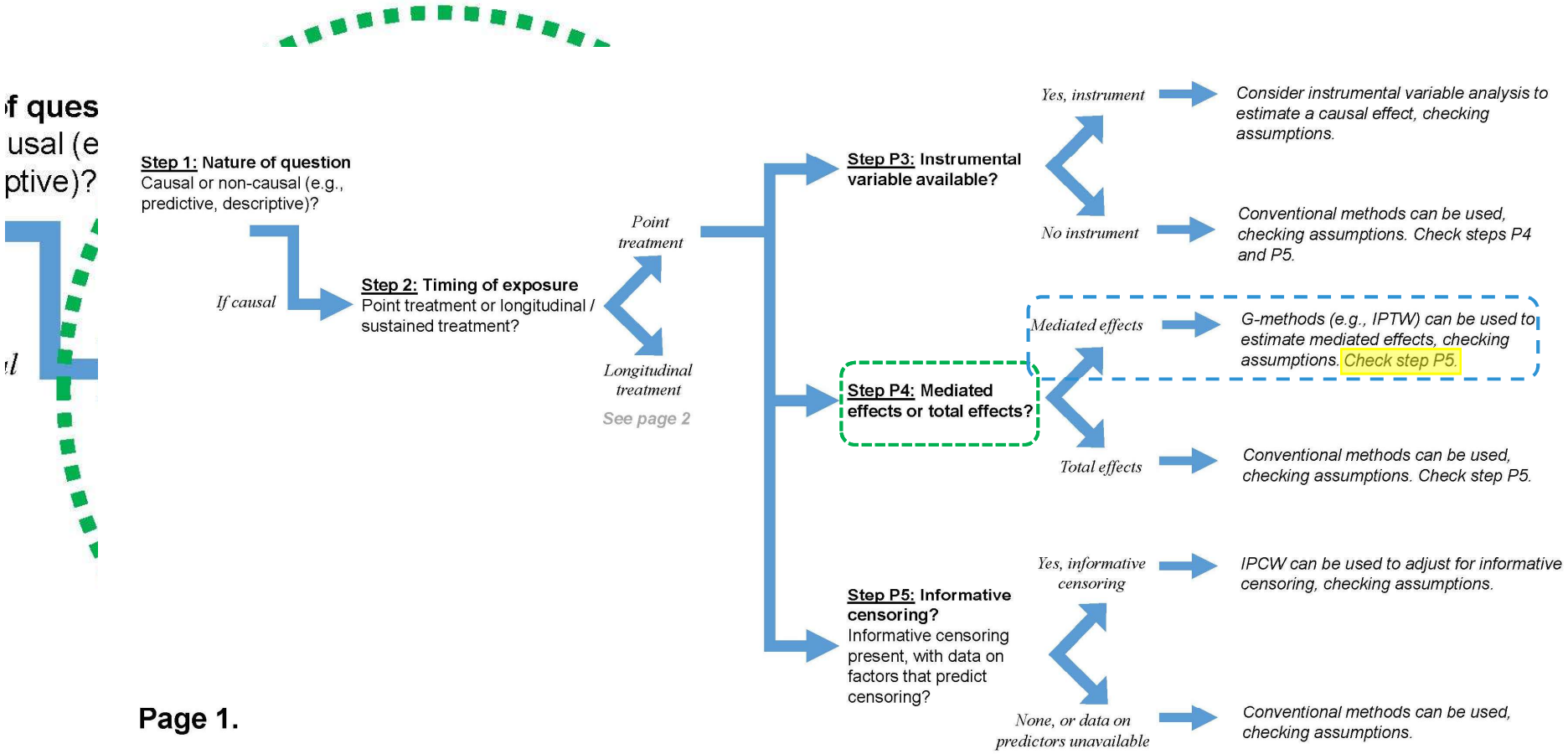
- This contrasts with conventional regression adjustment for confounders:

$$E(Y|A=a, \overline{W} = \overline{w}) = \beta_0 + \beta_1 * A + \overline{\beta_2} * \overline{W} + \varepsilon$$

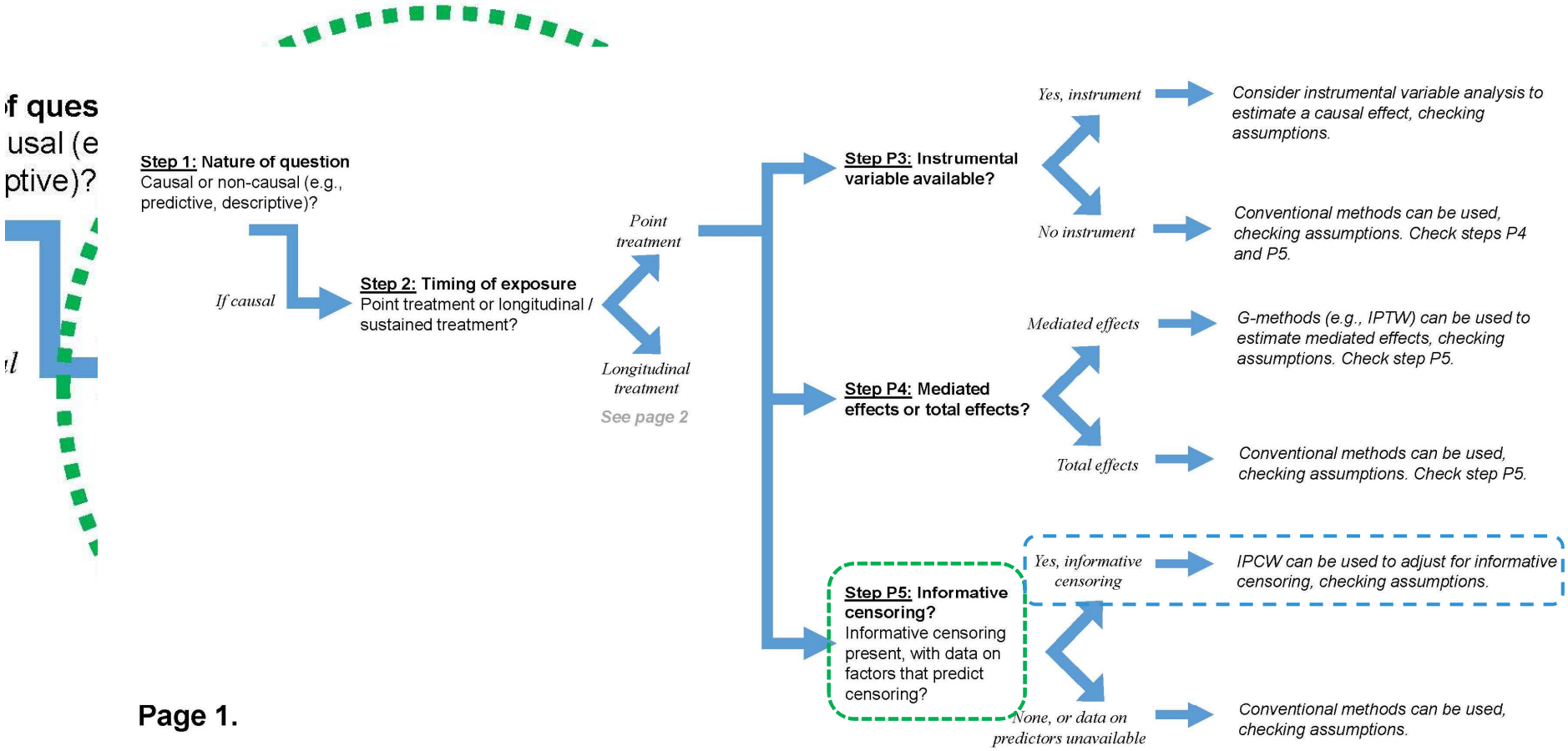
Agenda

- Recall motivation for using IPTW (decision tree)
- IPTW theory and mechanics
- Other applications of inverse probability weights (IPW):
  - Censoring/selection bias (IPCW)
  - Time-dependent confounding (IPTW)
- Data analysis activity

Recall our location on decision tree



Recall our location on decision tree

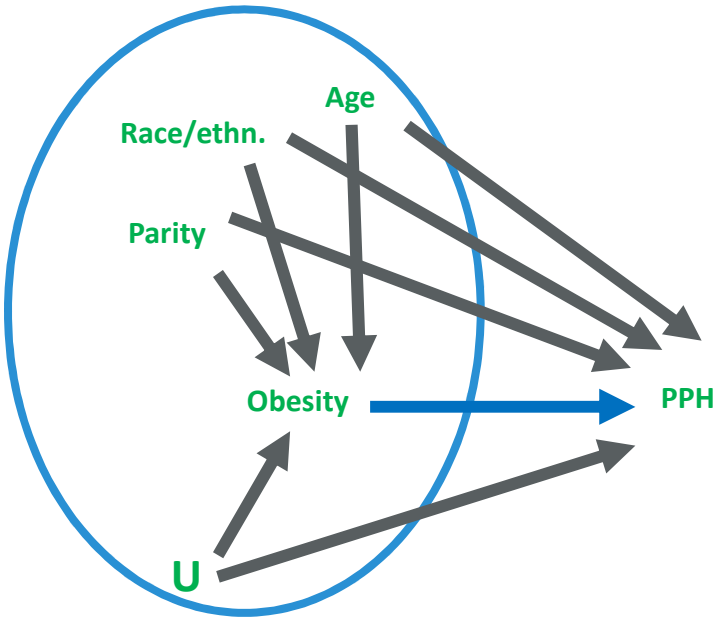


IPTW, IPCW, IPAW

- So far we have talked about IPW as applied to treatment: ***IPTW***
- The same technique can be applied to modeling the ***censoring mechanism*** (***S***).
- Then, ***IPCW*** can be used to adjust for censoring/selection bias.
- Both IPTW and IPCW can be combined (i.e., ***IPAW***, A=action).

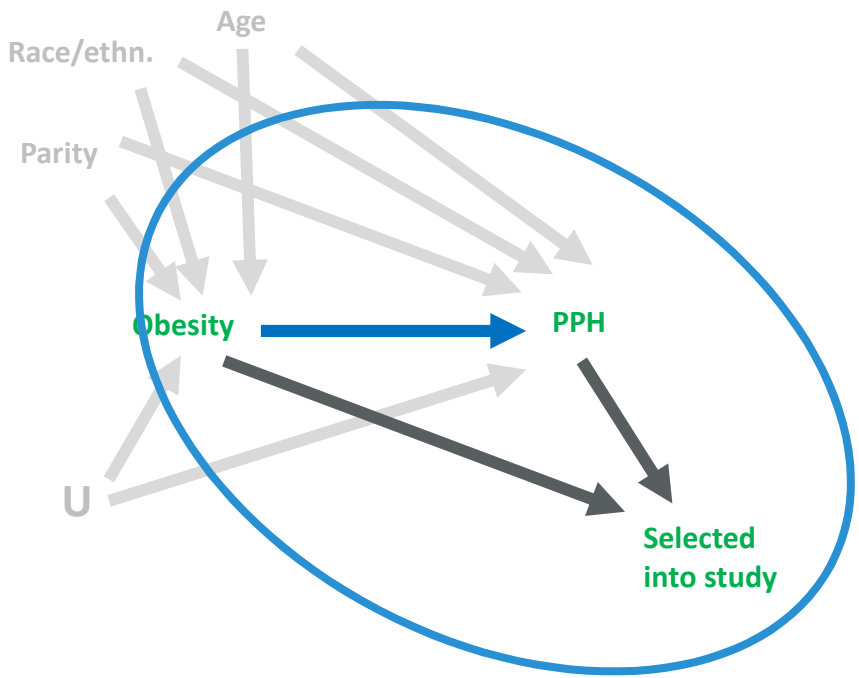
Modeling treatment vs. censoring

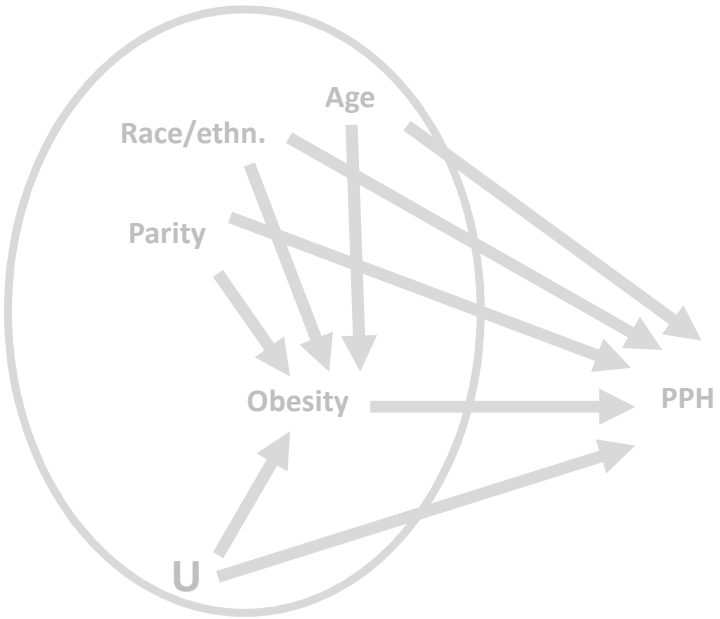
- Recall modeling the treatment mechanism using ***IPTW***.



Modeling treatment vs. censoring

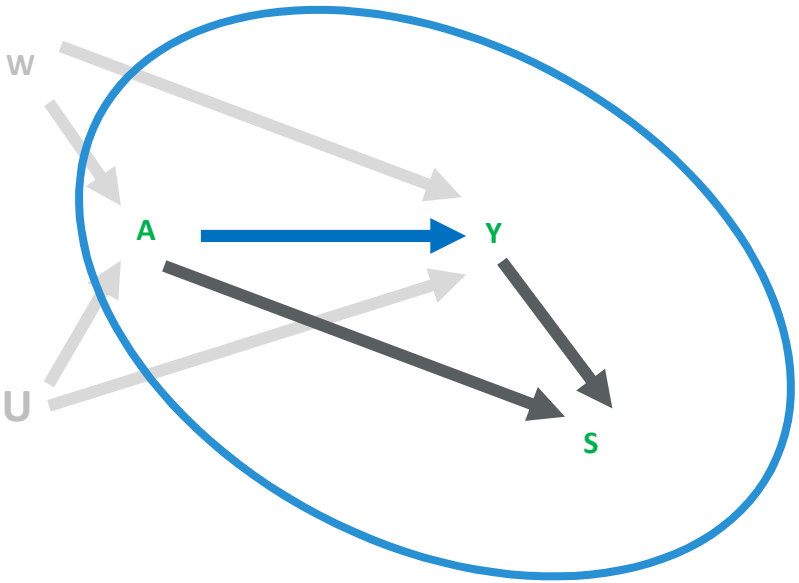
- Recall modeling the treatment mechanism using ***IPTW***.
- The same technique can be applied to modeling the ***censoring mechanism*** (denoted ***S***), using ***IPCW***.





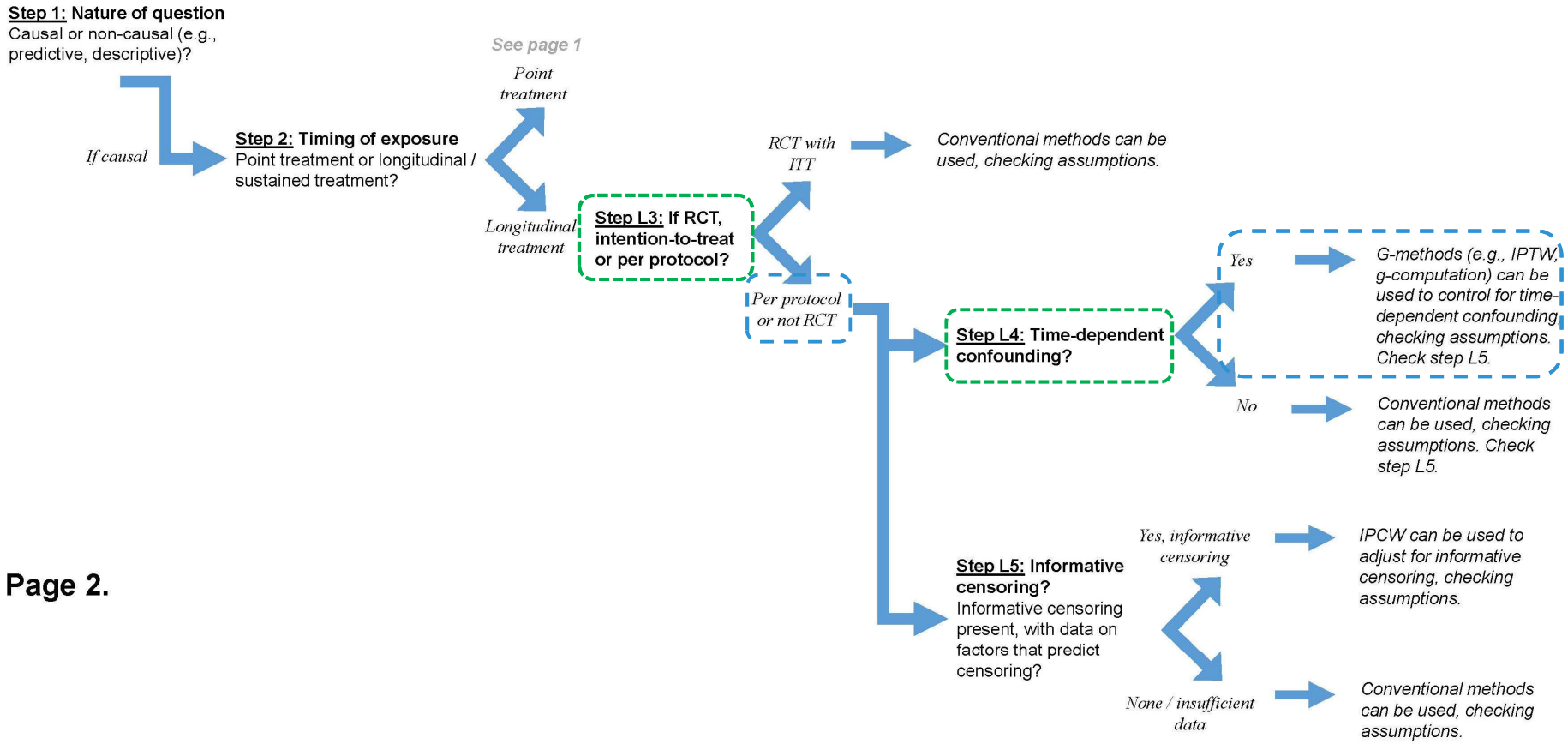
Censoring mechanism

- Suppose obesity exposure affects probability of participating in given prospective study: those with obesity are less likely to participate. (A→S)
- Suppose that people with postpartum hemorrhage (e.g., PPH) are also less likely to be enrolled. (Y→S)
- This results in ***selection bias*** (collider stratification).



- ***IPCW*** can be used to model the censoring mechanism and adjust for this selection bias.
- Instead of modeling Pr(*A)* weighting in IPTW, you model Pr(*S*) and weight in IPCW

Revisit decision tree: **time-dependent confounding of longitudinal tx.**

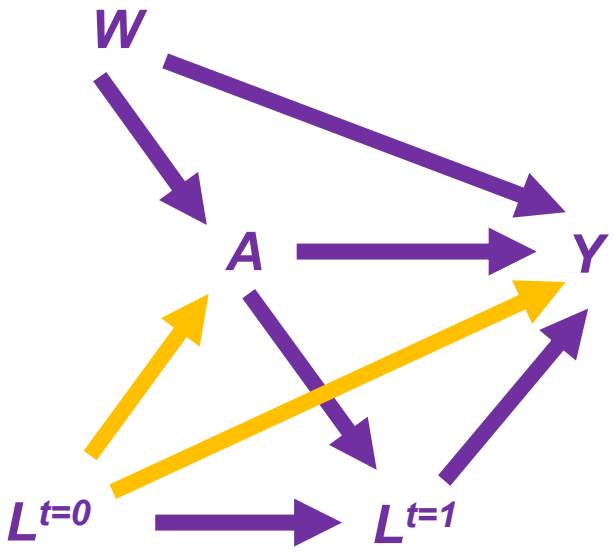


IPW application: time-dependent confounding

- Time-dependent confounding occurs when a time-varying confounder is both a confounder **and** a causal intermediate
- Also referred to as:
  - Treatment-confounder feedback
  - Time-varying confounding
  - Time-varying confounding affected by prior exposure
- More detailed descriptions in Snowden *JMWH* 2018, Naimi *IJE* 2017, Cole *AJE* 2008.
- IPTW can be used to adjust for this type of confounding.

IPW application: time-dependent confounding

- New notation:
  - Recall: ***W*** is a time-fixed confounder
  - $L^{T=t}$ for a time-varying confounder
  - ***t*** is the time-index (i.e., ***t***=0 is baseline; ***t***=1 is time-period 1; …)
- Time-dependent confounding occurs when a time-varying confounder ($L^{T=t}$) is a confounder ($L^{t=0}$) **and** a causal intermediate ($L^{t=1}$)
- Confounding pathway



Contrasts

- ***IPW***
  - **Marginal** interpretation
  - Modeling $\mathbf{E}(Y_a)$
  - Adjusts for confounders **without** conditioning.
  - **Can control for time-dependent confounding** w/o blocking causal pathway
  - Can estimate mediation effects in modern causal framework
- ***Regression adjustment***
  - **Conditional** interpretation
  - Modeling $\mathbf{E}(Y|A=a, \overline{W} = \overline{w})$
  - Adjusts for confounders **by** conditioning.
  - **Cannot control for time-dependent confounding** w/o blocking causal pathway

Agenda

- Recall motivation for using IPTW (decision tree)
- IPTW theory and mechanics
- Other applications of inverse probability weights (IPW):
  - Censoring/selection bias (IPCW)
  - Time-dependent confounding (IPTW)
- Data analysis activity

Data analysis activity

Inverse probability weighted logistic regression model for estimating direct / indirect effects.

Lange et al 2012[1]

Steps:

1) Specify model for the exposure
2) Specify model for the mediator
3) Determine weights
    1) Using an expanded dataset
4) Fit a generalized MSM for postpartum hemorrhage (Y) including only Obesity (A) and Obesity Star (A*) using an expanded dataset with repeating observations and weighted

[1]Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. American Journal of Epidemiology. 2012;176:190-5.

Step 1. Specify Exposure Model

$$Logit(Obesity|\bar{W})$$
$$= \beta_0 + \beta_1 \cdot age + \beta_2 \cdot race + \beta_3 \cdot educ. + \beta_4 \cdot parity$$

Step 2. Specify Mediator Model

$$Logit(Birth\ Mode|Obesity, \bar{W})$$

$$= \beta_0 + \beta_1 \cdot obesity + \beta_2 \cdot age + \beta_3 \cdot race + \beta_4 \cdot educ + \beta_5$$

$$\cdot parity + \beta_5 \cdot prior\ caeserean$$

Step 3: Determine Weights

$$W_i = \frac{P(A = A_i)}{P(A = A_i \mid C = C_i)} \frac{P(M = M_i \mid A = A_i^*, \ C = C_i)}{P(M = M_i \mid A = A_i, \ C = C_i)}$$

Step 3: Determine Weights:

- *First Fraction*

$$W_i = \frac{P(A = A_i)}{P(A = A_i \mid C = C_i)} \frac{P(M = M_i \mid A = A_i^*, \ C = C_i)}{P(M = M_i \mid A = A_i, \ C = C_i)}$$

- P is derived from the logistic regression of the exposure (A) on confounders (C).
- Standardized IPW

Step 3: Determine Weights:

- *Second Fraction*

$$W_i = \frac{P(A = A_i)}{P(A = A_i \mid C = C_i)} \frac{P(M = M_i \mid A = A_i^*, \ C = C_i)}{P(M = M_i \mid A = A_i, \ C = C_i)}$$

- P is derived from the logistic regression of the mediator (M) on exposure (A) and confounders (C).
  - Upweighting observations where the observed mediator value (M_i) would have been more likely to occur under a different exposure value (A_i^*) than the one actually observed (A_i).

Step 3: Determine Weights

- Second Fraction
- Repeat each observation
- *A* and M* are* auxiliary variables

Original data

ID	Obesity (A)	Caesarean (M)	PPH (Y)	Confound. (W)
1	1	0	0	1
2	0	0	1	1

Repeat data

ID	Obesity (A)	Obesity (A*)	Caesarean (M)	Caesarean (M*)	PPH (Y)	Confound. (W)
1	1	1	0	0	0	1
1	1	0	0	--	0	1
2	0	0	0	0	1	1
2	0	1	0	--	1	1

Step 3: Determine Weights

$$W_i = \frac{P(A = A_i)}{P(A = A_i \mid C = C_i)} \frac{P(M = M_i \mid A = A_i^*, \ C = C_i)}{P(M = M_i \mid A = A_i, \ C = C_i)}$$

Step 4. Fit a generalized MSM

3) Fit a generalized MSM for PPH (Y) including only Obesity (A), and Obesity Star (A*) using an expanded dataset with repeating observations and weighted

$$logit\left(E\left[Y_{a^1,M_{a^2}}\right]\right) = \beta'_0 + \beta'_1 A^1 + \beta'_2 A^2 + \beta'_3 (A^1 \cdot A^2)$$

weights = IPW

Calculate Percent Mediated

Contrast Estimate Results						
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error*	Alpha
NDE	0.5978	0.5710	0.6240	0.3962	0.0564	0.05
OR^{NDE}				1.4862	0.0838	0.05
NIE	0.5242	0.5197	0.5287	0.0969	0.0091	0.05
OR^{NIE}				1.1017	0.0101	0.05

Total Effect = $OR^{NDE} * ORNIE$ = 1.637

For OR, Percent Mediated = $\frac{OR^{NDE}.(OR^{NIE}-1)}{OR^{NDE}.OR^{NIE}-1}$ = 0.237 or 24%

*Robust standard errors are not valid here – Standard errors need to be obtained by bootstrapping.

Thank you!