Workshop timeline

8:00 – 8:50 … Topic: Overview of theory and applications

8:50 – 9:00 … Break

9:00 – 9:30 … Applied example

9:30 – 9:40 … Break and code prep

9:40 – 11:15 … Applied session

11:15 – 11:30 … Summary and wrap-up

Hypothetical interventions for exposure mixtures

Practical theory and applications for epidemiologists

PART I: Theory

Causal inference and mixtures

- One major barrier to causal inference in mixtures is that none of the simple, textbook examples apply
- Today we'll create a map from the things you may know (textbook causality - with review) to things that I hope you can become more comfortable with (causality in mixtures)
- You can access materials at https://github.com/alexpkeil1/ISEE2025_Mixtures_Workshop, so I will move fast, but please ask questions!

Scope

- The discussion today is in restricted to a very specific question that might be asked of mixtures data (use your own definition)
- "What would be the health impact on some health outcome if we could modify some or all of the exposures in the mixture?"[1]
- I prefer to call this endeavor "causal effect estimation"[2], but the label "causal inference" persists, so I use it

---

[1]I may speak as though this is the only useful question, but I do not believe that.

[2]Greenland, S. (2017). For and against methodologies: some perspectives on recent causal and statistical inference debates. European journal of epidemiology, 32(1), 3-20.

Causal inference for the extremely busy

- Causal inference combines (sometimes unverifiable) assumptions and past observations to sharpen knowledge about how we can change the future or what we should have done differently in the past

- Causal inference is not the application of special methods that distinguish causation from correlation in a given data set

- Tools of causal inference are effective for evaluating new methods re: do they answer the question I want to ask?

What is a cause?

“Cause” has many potential definitions

What is a cause?

Many definitions (or at least common uses) are incomplete[1], ambiguous[2] or too restrictive[3]

---

[1]e.g. Granger causality
[2]i.e. we cannot map them to precise mathematical statements (e.g. Bradford Hill)
[3]e.g. deterministic causality

What is a cause?

A goal of improving (and not just observing) public health leads to a policy[1] definition of a cause

[1]A “policy” is a function that takes as input a current state and outputs an action. This definition comports with common usage of “public health policy” and so I use in preference to “regime.”

Policy definition of a cause[1]

Exposure causes an outcome if a manipulation[2] of the exposure (e.g. via intervention) would change the outcome[3]

---

[1]e.g. a policy could include random assignment to treatment/control arms

[2]Robins and Greenland (2000) J. Am. Stat. Assoc.

[3]This is packed and will be formalized

Policy and decisions

Causal effects contrast one policy with another[1]

Causal effect estimation allows “optimal” policy choices

Epidemiologic causal inference is about improving decisions for groups in context

---

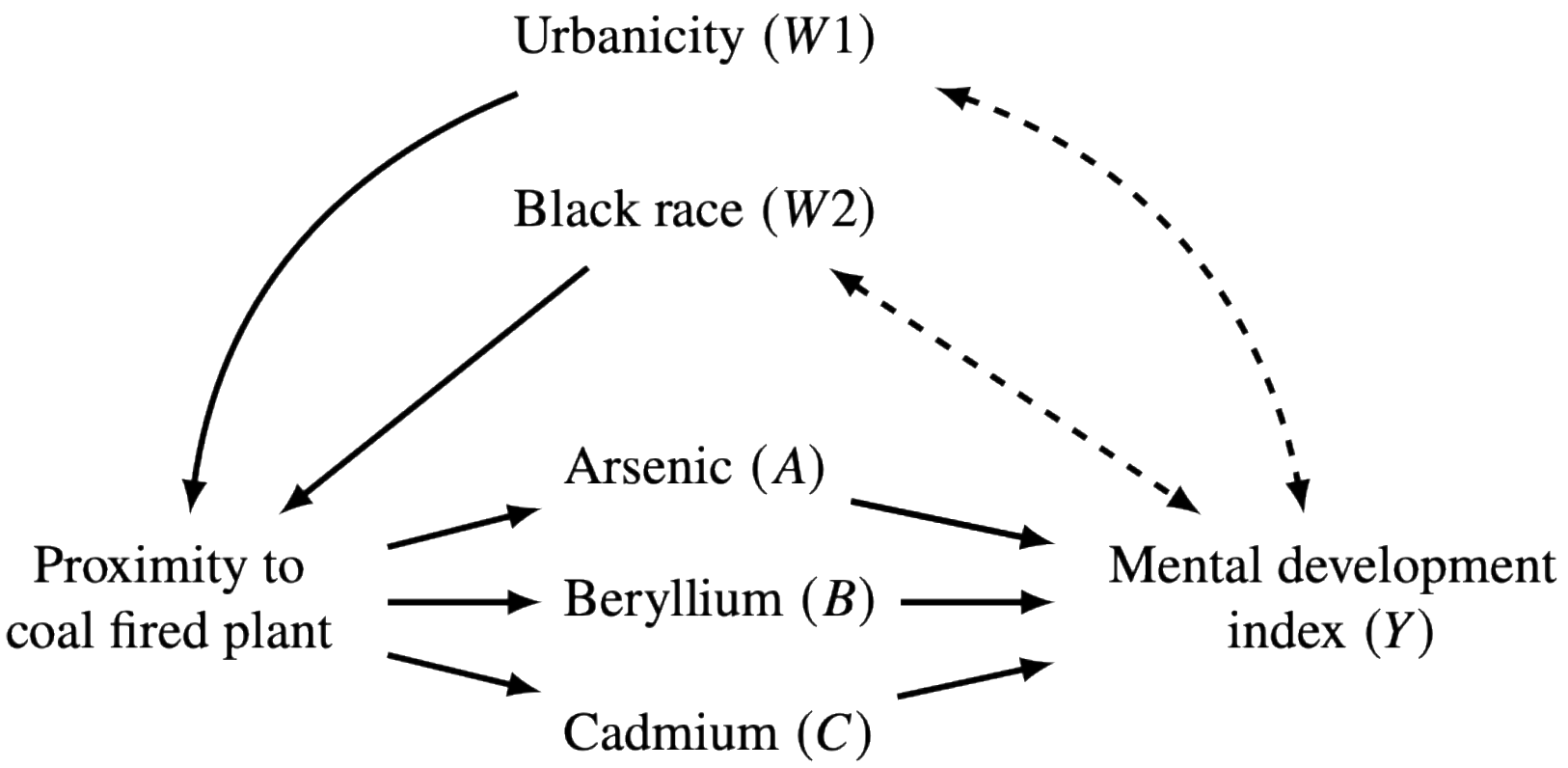
[1]Or define a function in the case of continuous policies like dose-responses

Motivating example: coal fired power plants and cognitive development

- Burning coal for energy produces many byproducts, including a mixture of air toxics
- Some of these have known detrimental effects on early-life cognitive outcomes
- Closure of coal-fired plants has been associated with improvements in cognitive outcomes
- (So that I can share the data) I performed a simulation study of exposure to coal-fired power plant emissions and mental development index

Motivating example: coal fired power plants and cognitive development

- Simulated N=2020 3 yr olds
- Annual mean residential air exposures (ug/m3)
- W1, W2 associated with Y via unmeasured, historical factors (social determinants)
- Coal plants near urban centers, and race associated with proximity to coal plants via historical factors (e.g. redlining)



Motivating example: coal fired power plants and cognitive development

(Possible) Causal questions of interest:

Causal independent exposure-response How does the population average MDI change as we increase Arsenic but hold other exposures constant

Causal joint exposure-response How does the population average MDI change as we increase all exposures by the same amount

Causal attributable difference How does the population average MDI change after we eliminate all/some exposures?

Causal generalized impact difference How does the population average MDI change after we reduce all/some exposures by some policy-relevant amount?

A note on study questions

- Causal inference allows us to ask questions with answers that make sense to people who don’t know what regression is, without sacrificing rigor

- I always encourage students to ask causal questions because it helps focus and choose methods even when causality is hopeless.

Foundational concepts

- Counterfactuals
- Potential outcomes and consistency
- Causal identification assumptions

Notation, notation notes

Univariate exposure/treatment A, with realized values a (binary for now)

Univariate outcome $\boldsymbol{Y}$**,** $\boldsymbol{y}$

Confounders W , w

Non-confounding covariates V , v

Individuals Individuals will be noted with i subscripts, as little as necessary (e.g. y_i)

Simplification Except when necessary, I will introduce most concepts for time-fixed data

Counterfactuals

Counterfactuals refer to the policies that are counterfactual for each individual: unseen doppelgangers that follow each policy of interest (here “yes” or “no” exposure)

id	W	A^1	Y	
1	2.9	yes	no	factual
1_y	2.9	yes	?	counterfactual
1_n	2.9	no	?	counterfactual
2	0.5	no	yes	factual
2_y	0.5	yes	?	counterfactual
2_n	0.5	no	?	counterfactual

[1]Here, I distinguish between "setting" *A* to some value via a policy versus "observing" *A* at the same value

Potential outcomes

Potential outcome: Y^a, or the value[1] of the outcome Y we would have observed, had exposure been set to some value a

[1]These are "deterministic" potential outcomes. We could similarly define "stochastic" potential outcomes via a probability distribution on Y^a.

Potential outcomes and causal effects

An <u>individual</u> (additive) causal effect of "yes" vs. "no" exposure is defined as

$$Y_i^{yes} - Y_i^{no}$$

In this sense, **causal effects are defined without data**, so to learn anything about causal effects we need a way to link what we see (data and/or priors), with what could be (potential outcomes).

---

[1]In the potential outcomes framework, this is a way of saying "no causation without manipulation" - causal effects are undefined without hypothetical manipulation.

Potential outcomes and causal consistency

The link from our factual world to counterfactual observations happens via causal consistency, which is that $Y^a = Y_i$ if $A_i = a$. That is, we would expect an individual to have identical outcomes if we set $A = a$ via policy or if we observed $A = a$.[1]

[1] There is a second assumption here that will be discussed later, which is included in the alternative stable unit treatment value assumption (SUTVA) that was originally used to link counterfactuals with observed data

Potential outcomes and causal consistency

Given causal consistency, we can fill in some of the table

id	W	A	Y	
1	2.9	yes	no	factual
1_y	2.9	yes	?	counterfactual
1_n	2.9	no	no	counterfactual
2	0.5	no	yes	factual
2_y	0.5	yes	?	counterfactual
2_n	0.5	no	yes	counterfactual

Potential outcomes and causal consistency

Here's another way you may see it[1]:

id	W	A	Y	Y^{yes}	Y^{no}
1	2.9	yes	no	no	?
2	0.5	no	yes	?	yes

[1]Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688–701.

The fundamental problem of causal inference

Recall that we define an individual, additive[1] causal effect as the difference $Y_i^{yes} - Y_i^{no}$

id	W	A	Y	Y^{yes}	Y^{no}	$Y^{yes}-Y^{no}$
1	2.9	yes	no	no	?	?
2	0.5	no	yes	?	yes	?

Even with causal consistency, at least one of the potential outcomes necessary for defining a causal effect will always be missing. This is known as the **fundamental problem of causal inference**. To make progress we need more assumptions.

[1]We could also use ratios and the relative scale.

Individual and average causal effects

Even if we can't estimate individual causal effects, we can use additional assumptions to estimate average potential outcomes (***E*** $(Y^a \mid W = w)$), and hence average causal effects:

- **Population/sample average causal effect** $E(Y^a) - E(Y^{a*}) = E(Y^a - Y^{a*})$
- **Average affect among the "a-treated"** $E(Y^a - Y^{a*} \mid A = a)$
- **Causal dose-response** $f(y^a)$
- **Conditional average causal effects** $E(Y^a - Y^{a*} \mid W = w)$

Whose average?

The definition of a population average causal effect presupposes a specific population, often called the **target population**.

Whose average?

Defining the target population is essential to evaluating the utility of estimates of causal effects, and often necessitates concepts of **generalizability** or **transportability**

Whose average?

Formally, **generalizing** or **transporting** causal effects relies applying conditional average causal effects (conditional on ***W*** and/or ***V***) to populations with different distributions of ***W*** and/or ***V*** from the study sample.

Whose average?

W and ***V*** are sometimes referred to as the “context", or “state”, since they are often non-modifiable factors that can influence the effectiveness of a policy. We may think of factors like exposure susceptibility, race, and socioeconomic status as classical examples of factors that differ across populations and by which conditional causal effects may vary greatly.

Whose average?

The term “**average causal effect**” is thus meaningless without first defining the target population, which often is not the study population.

Causal identification conditions

Returning to the main thread, causal consistency alone is not sufficient to license the estimation of average causal effects.

Causal identification conditions

no interference

Consider if ***A*** is vaccination and ***Y*** is SARS-COV-2 infection. Your infection, if not vaccinated (Y^{no}) doesn't just depend on your vaccination status, but also the vaccination status of your neighbor.

Causal identification conditions

no interference

If one's potential outcome depends also on others' exposure the potential outcome would have to be denoted by your exposure, as well as everyone else's exposure, denoted by the hideous looking expression

$$Y^{a_i,\{a_j \in A: j \neq i\}}$$

This makes the math much more difficult.[1] Under the no interference assumption (based on subject matter knowledge), your potential outcomes don't depend on the exposures of others, so we are safe to just use simpler notation and math.

---

[1] though causal inference is sometimes still possible: Hudgens et al (2008) *J Am Stat Assoc*

Causal identification conditions

no interference

Causal consistency and **no interference** give us "observed" potential outcomes, where we know your potential outcome under the exposure that you, in fact, received. However, recall that causal inference requires us to know something about the potential outcomes under counterfactual exposures/policies, as well.

Causal identification conditions

no interference, conditional exchangeability

The link to counterfactual policies is provided by **conditional exchangeability**[1], given by

$$Y^a \perp A | W = w$$

which reads as "The potential outcome under a given policy is independent of exposure, given confounders." This assumption means that, in a stratum of confounders ***W*** , we can consider the "observed" potential outcomes to stand in for the "missing" potential outcomes because in that stratum the individuals are "exchangeable" with each other

---

[1]This is sometimes referred to as a “no unmeasured confounding or selection bias” assumption

Causal identification conditions

no interference, conditional exchangeability

$$Y^a \perp A | W = w$$

Conditional exchangeability doesn't give us individual potential outcomes but we can "impute" **average potential outcomes in strata of confounders** via:

$$E(Y^a | A \neq a, W = W) = E(Y^a | A = a, W = W)$$

While not obvious, this has *almost* given us enough to estimate average potential outcomes for an entire population, and thus causal effects.

Causal identification conditions

no interference, conditional exchangeability, positivity

- Notably, the conditional causal effect

$$E(Y^{a^*} - Y^a | W = w) = E(Y | A = a^*, W = w) - E(Y | A = a, W = w)$$

only makes sense if it is possible to observe the combination ***A = a, W = w***. One way to re-write this is:

$$Pr(A = a | W = w) > 0$$

for each policy a being compared. This assumption is referred to as **positivity**

Causal identification conditions

no interference, conditional exchangeability, positivity

Aside: Positivity means that, for a given policy a, that value of exposure must be possible at all joint levels of confounders. **Sparsity**[1] can occur when $\boldsymbol{A = a}$ is possible, but simply not observed in the data.

Here's a distinction:

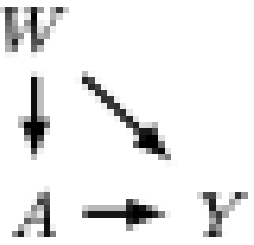
Non-positivity: always biased, and our causal question may be ill-posed (e.g. effect of hysterectomies among people born without a uterus)

Sparsity: biased, but we may reduce/eliminate bias with more data or a model

---

[1]this has been called "stochastic non-positivity"

G-computation



Task: get average potential outcome at $A = 0$ (Y^0 under no interference)

$$Pr\left(Y^0 = 1\right) = \sum_w Pr\left(Y^0 = 1 | W = w\right) Pr\left(W = w\right) \qquad \text{Law of total probability}$$

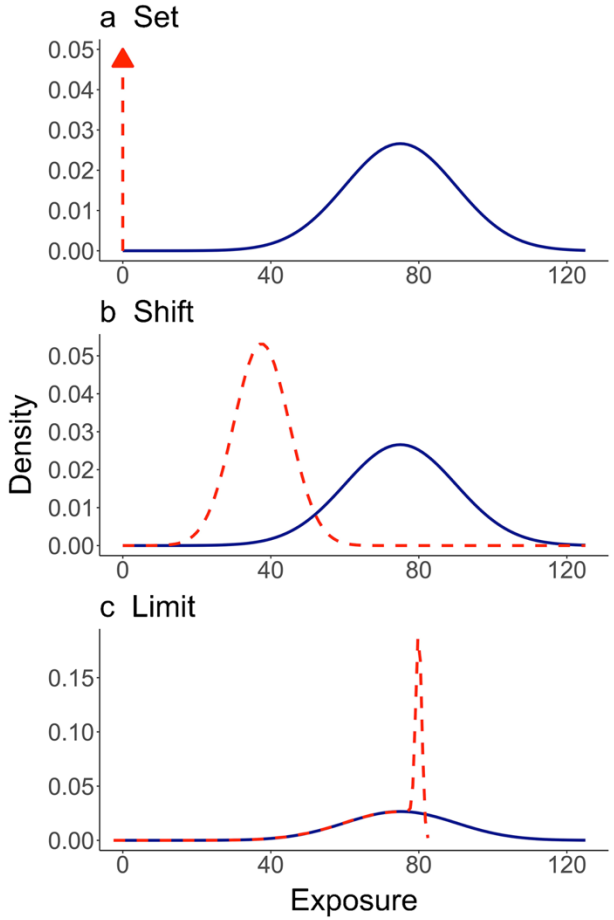
$$= \sum_w Pr\left(Y^0 = 1 | A = 0, W = w\right) Pr\left(W = w\right) \qquad \text{Exchangeability}$$

$$= \sum_w Pr\left(Y = 1 | A = 0, W = w\right) Pr\left(W = w\right) \qquad \text{Consistency, positivity}^1$$

The final step shows how we can do this in data: 1) fit a standard regression model for Y and 2) average the predictions over the entire dataset (easy!)

G-computation – many different policy effects

causal inference for continuous exposures



- We can set all individuals to one level ***A=a***
- We can also have an intervention where the intervention can set individuals to different values (e.g. drawing from a distribution)

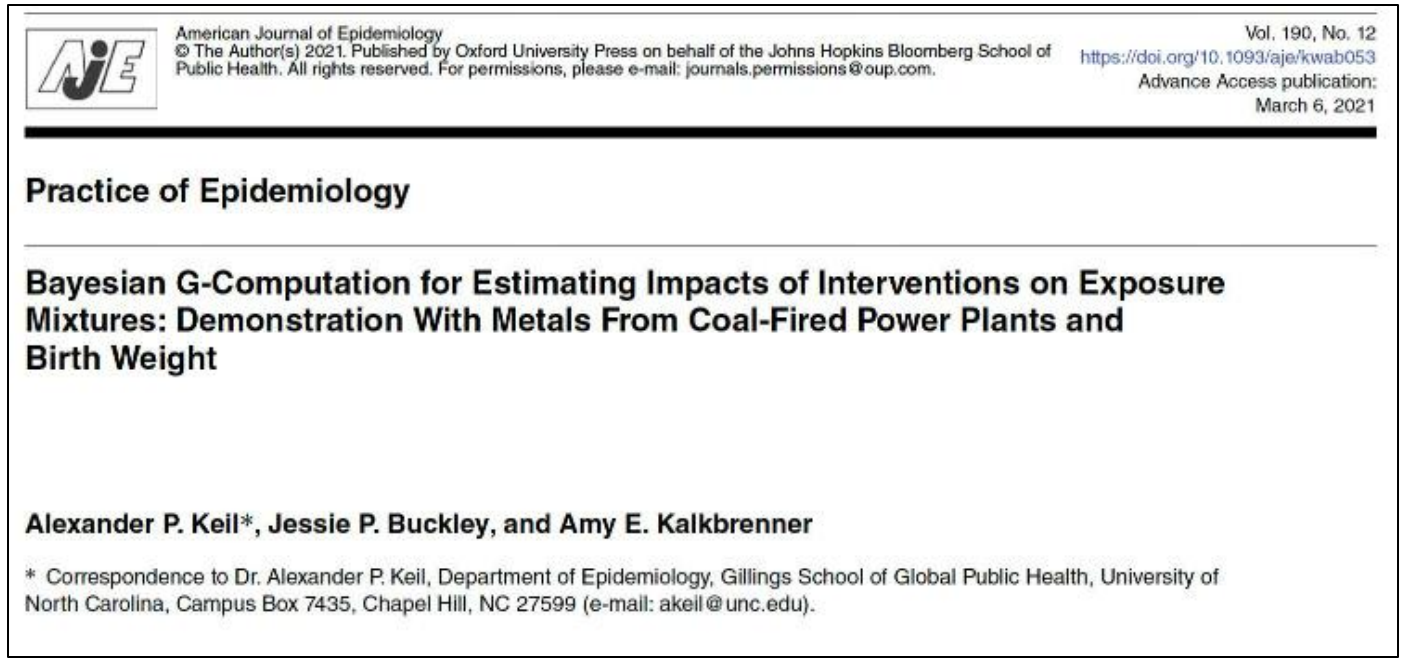
$$Pr(Y_g = 1) = \sum_w Pr(Y = 1|g \sim f(w)|W = w)Pr(W = w)$$

- Or “capping” the observed exposures at a specific value
- And it is a trivial extension to consider more than one exposure

Examples with mixtures

- Hypothetical example in R with coal fired power plant emissions and cognitive development: https://github.com/alexpkeil1/ISEE_2020_causal/tree/main/analyses/code

- Real example with coal fired power plant emissions and birth weight (Keil et al 2020)



American Journal of Epidemiology

Vol. 190, No. 12
https://doi.org/10.1093/aje/kwab053
Advance Access publication:
March 6, 2021

Practice of Epidemiology

Bayesian G-Computation for Estimating Impacts of Interventions on Exposure Mixtures: Demonstration With Metals From Coal-Fired Power Plants and Birth Weight

Alexander P. Keil*, Jessie P. Buckley, and Amy E. Kalkbrenner

* Correspondence to Dr. Alexander P. Keil, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Campus Box 7435, Chapel Hill, NC 27599 (e-mail: akeil@unc.edu).