Causal inference for exposure mixtures

ISEE 2020 virtual pre-conference workshop → ISEE sponsored webinar

Alexander Keil
e: akeil@unc.edu
: @PronouncedKeil

January 7, 2021

Historical note

Yesterday (2020-01-06) armed insurrectionists, at the suggestion of the sitting president of the US, broke through police barricades and entered the US Capitol building in Washington DC. The riot was timed to coincide with the proceedings at which the peaceful transfer of executive power was to occur. As of now, 14 DC Metropolitan police officers were injured during the riot, and 4 individuals died including 1 as a result of a gunshot wound inside the Capitol building. Multiple pipe bombs and molotov cocktails were found in the subsequent cleanup. Metropolitan police officers were shown, on camera, explicitly allowing some rioters into the Capitol, and assisting rioters as they left the building, without arresting them. The sitting US president and members of the US legislative branch continue to openly assert, against all evidence, that the incoming US president was elected fraudulently and are openly advocating for selective disenfranchisement of citizens of their country and denying, against evidence, the role of their supporters in the riots.

This talk will start easy, then get difficult, and then get easier

materials:

https://github.com/alexpkeil1/ISEE_2020_causal/archive/main.zip

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)

This session will be recorded, and you can access materials, 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 - forgive my enthusiasm - 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[3], ambiguous[4] or too restrictive[5]

[3] e.g. Granger causality
[4] i.e. we cannot map them to precise mathematical statements
[5] e.g. deterministic causality

What is a cause?

A goal of improving (and not just observing) public health leads to a **policy**[3] definition of a cause

[3] 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[6]

Exposure **causes** an outcome if a manipulation[4] of the exposure (e.g. via intervention) would change the outcome[5]

[4] e.g. a policy could include random assignment to treatment/control arms
[5] Robins and Greenland (2000) *J. Am. Stat. Assoc.*
[6] This is packed and will be formalized

Policy and decisions

- Causal effects contrast one policy with another[7]

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

Policy and decisions

- Causal effects contrast one policy with another[7]
- Causal effect estimation allows “optimal” policy choices

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

Policy and decisions

- Causal effects contrast one policy with another[7]
- Causal effect estimation allows “optimal” policy choices
- Epidemiologic causal inference is about improving **decisions** for groups in context

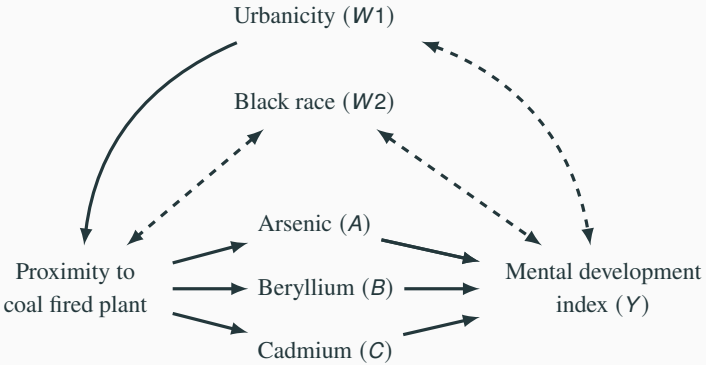
[7] Or define a function in the case of continuous policies like dose-responses

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/m^3)
- $W1, W2$ associated with Y via unmeasured, historical factors (education, racism)
- Coal plants near urban centers, and race associated with proximity to coal plants via 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 Y, y

Confounders $\boldsymbol{W}$, $\boldsymbol{w}$

Non-confounding covariates $\boldsymbol{V}$, $\boldsymbol{v}$

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 dopplegangers that follow each policy of interest (here “yes” or “no” exposure)

id	W	A[8]	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

[8]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[9] of the outcome Y we would have observed, had exposure been set to some value a

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

Potential outcomes

An individual (additive) causal effect of "yes" vs. "no" exposure is defined as $Y_i^{yes} - Y_i^{no}$.[9]

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).

[9] 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 $Y_i^a = Y_i$ 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 $A = a$.[10]

[10]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^{yes}	2.9	yes	no	counterfactual
1^{no}	2.9	no	?	counterfactual
2	0.5	no	yes	factual
2^{yes}	0.5	yes	?	counterfactual
2^{no}	0.5	no	yes	counterfactual

Potential outcomes and causal consistency

Here's another way you may see it:[10]

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

[10] 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[11] 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	?

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

The fundamental problem of causal inference

Recall that we define an individual, additive[11] 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.

[11] 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 ($E(Y^a|\boldsymbol{W} = \boldsymbol{w})$), 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" $E(Y^a - Y^{a^*}|A = a)$

Causal dose-response $f(y^a)$

Conditional average causal effects $E(Y^a - Y^{a^*}|\boldsymbol{W} = \boldsymbol{w})$

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**[12]

[12] Ask Daniel Westreich: westreic@email.unc.edu

Whose average?

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

Whose average?

$\boldsymbol{W}$ and $\boldsymbol{V}$ 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 vacuous 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

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

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: (the hideous looking $Y^{a_i, \{a_j \in A : j \neq i\}}$).

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

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: (the hideous looking $Y^{a_i, \{a_j \in A: j \neq i\}}$). This makes the math much more difficult.[12] 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.

[12] 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**[12], given by

$$Y^a \amalg A|\mathbf{W} = \mathbf{w}$$

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 $\boldsymbol{W}$, 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

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

Causal identification conditions No interference, conditional exchangeability

$$Y^a \perp\!\!\!\perp A | \mathbf{W} = \mathbf{w}$$

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

$$E(Y^a | A \neq a, \mathbf{W} = \mathbf{w}) = E(Y^a | A = a, \mathbf{W} = \mathbf{w})$$

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

Notably, the conditional causal effect

$$E(Y^a|A = a, \mathbf{W} = \mathbf{w})$$

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

$$Pr(A = a|\mathbf{W} = \mathbf{w}) > 0$$

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

Aside: Positivity means that, for a given policy a, that value of exposure must be *possible* at all joint levels of confounders. **Sparsity**[12] can occur when $A = a$ 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

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

Causal identification conditions

No interference, conditional exchangeability, positivity

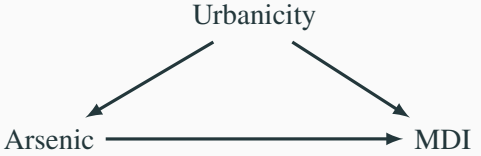
No interference, conditional exchangeability and, positivity are known as *causal identification conditions* because they are sufficient to "identify" a causal effect from observed data.[12] In a perfectly run randomized trial, these conditions will hold by design. In observational studies, we must rely on subject matter knowledge to judge how well these assumptions are met.

[12] but not necessarily the data in hand

Conditional exchangeability and DAGs

Conditional exchangeability can be (roughly) read off of a causal directed acyclic graph (DAG). These two tools inhabit different causal inference paradigms (potential outcomes versus do-calculus[13]), but they are mathematically equivalent. For example, the following conditional exchangeability statement (among others) is consistent with the following DAG.

$$MDI^{no\ arsenic} \amalg Arsenic | Urbanicity = yes \quad (1)$$



This is another place where "subject matter knowledge" is essential for causal inference - if we don't know our DAG we can't assess causal assumptions.

[13] Pearl, Judea. Causality. Cambridge University Press, 2009.

Minimal causal unit: conditional effects

Together the causal identification conditions give us the following:

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

Which we can read as: the average *potential* outcome under the policy "Set A=a" in a group where $W = w$ is equal to the average *observed* outcome in a subset of your data where $A = a$ and $W = w$

This link allows us to "see" causal effects in actual data. If we don't believe we can get a conditional effect somewhere, we generally can't do causal effect estimation.

So you are telling me we need to adjust for confounders?

Potential outcomes/causal inference has been criticized for being a re-statement what we already know: we need to adjust for confounders.

While it's true that causal inference has formalized the definition of confounding (via DAGs)

Extending foundational concepts with our coal-fired power plant example

- Why textbook examples fail mixtures
- Inverse probability weighting and G-computation
- Wrap-up

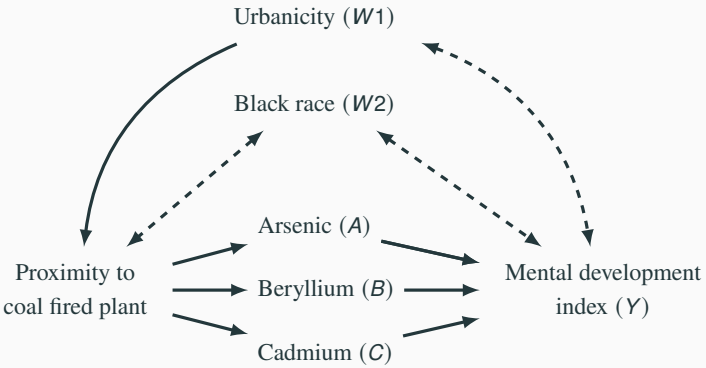
Textbook examples fail us for mixtures

Our textbook examples have only included on single binary exposure A.

But mixtures have continuous, and potentially highly exposures, so we'll almost always have sparsity where certain ranges of exposure just are not observed for different covariate/exposure strata.

And a mixture of multiple exposures (Arsenic, Beryllium, Cadmium) means we can't rely (without modification) on the wealth of methods for single exposures.

We'll investigate these issues without resorting to (too much) notation. Recall our (simulated) example



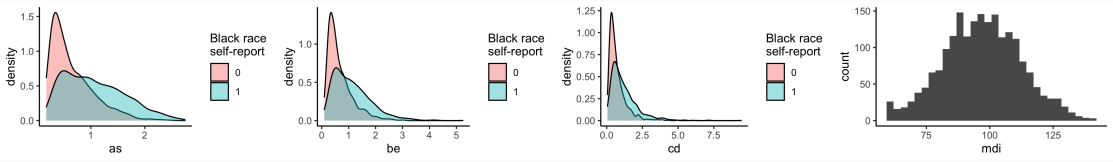
Basics of the data

The first 5 observations

urbanicity	black	as	be	cd	mdi
1	0	0.814	0.931	1.64	82.5
0	0	0.243	0.354	0.243	93.9
1	0	0.404	0.406	0.665	123
0	1	0.366	0.241	0.218	115
1	0	0.245	0.347	0.335	110

Pearson correlation matrix

	as	be	cd
as	1		
be	0.86	1	
cd	0.76	0.68	1



Textbook examples and continuous exposures

Under our DAG, for a single exposure (say arsenic), the "minimal causal unit" that we need is:

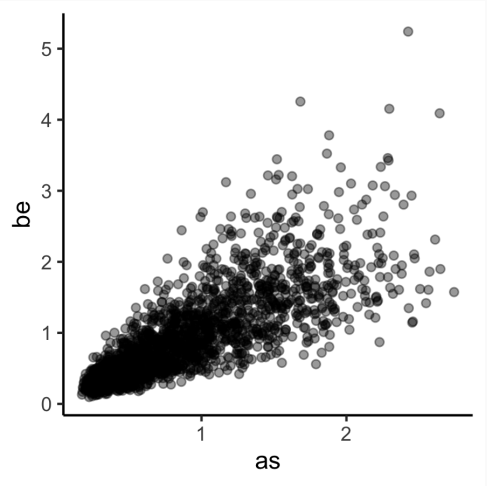
$$E(MDI|A = a, B = b, C = c, Confounders)$$

Roughly, to identify a causal effect of increasing arsenic by 1 unit (e.g. an independent causal dose response), we would need to observe, roughly, that at all values of the other exposures and confounders, we see some variation in arsenic of about 1 unit.[14]

[14] Sort of, this is language for discrete variables, which is fuzzy to apply here but to be more exact requires measure theory :(

Textbook examples and continuous exposures

Visually, we can see whether sparsity is an issue (it is, and we've known it for a long time)[15]

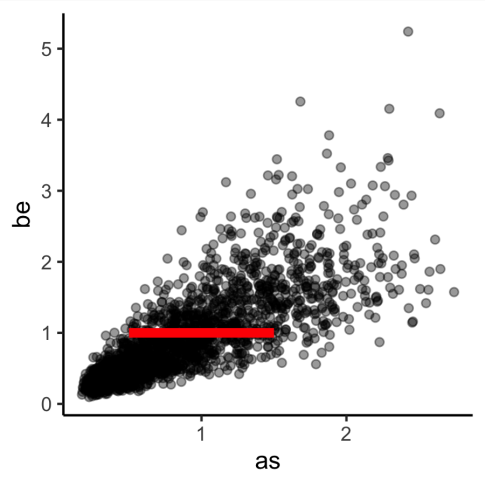


[15]Snowden, et al. Framing air pollution epidemiology in terms of population interventions, with applications to multi-pollutant modeling. Epidemiology 26.2 (2015): 271.

Textbook examples and continuous exposures

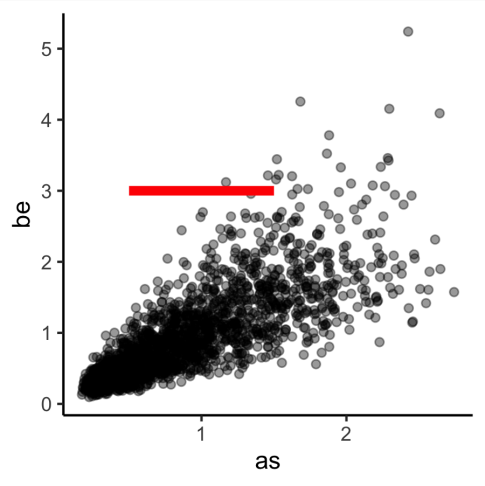
Here we can see that a one unit change (1.5 $\rightarrow$ 2.5) in arsenic is possible and (informally) identified when Beryllium = 1.0 ug/m^3

"Identification" here just informally means that we have some data points on both the left and right sides of the red bar, where the MDI for those individuals would be used in estimating the causal contrast (very roughly speaking)



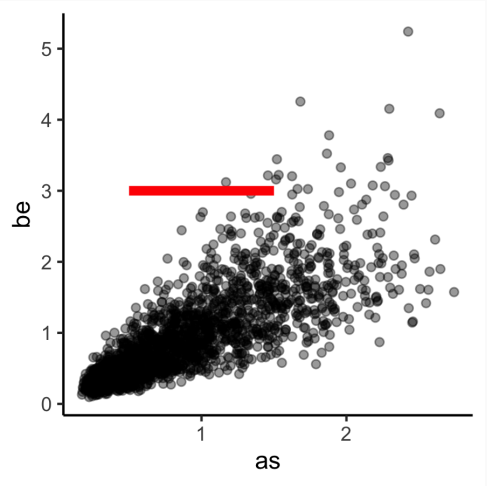
Textbook examples and continuous exposures

It may be possible (positivity) but it is not observed (sparsity) when Beryllium = 3.0 ug/m^3



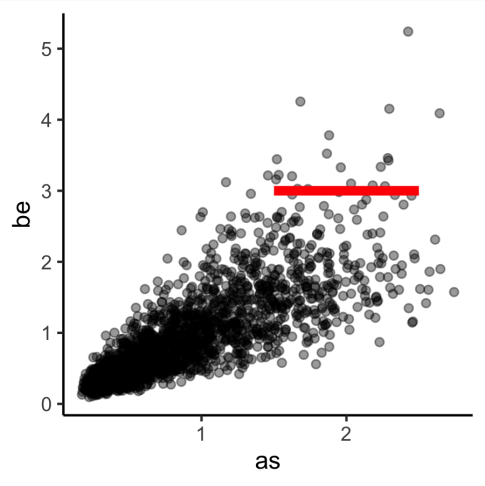
Textbook examples and continuous exposures

This issue is well described and it seems hopeless that we'll be biased.



Textbook examples and continuous exposures

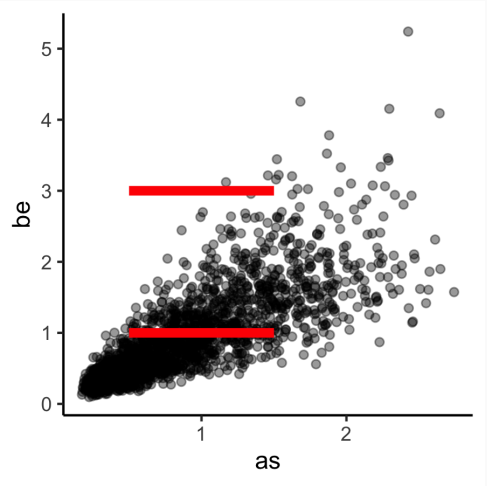
However, note that, a 1 unit change (2.5 → 3.5) in arsenic is observed when Beryllium = 3.0 ug/m^3



Textbook examples and continuous exposures

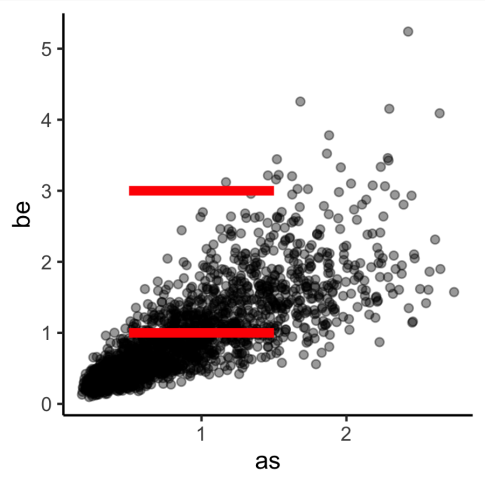
If we can assume that the rate of change (e.g. regression coefficient) of the outcome is modeled accurately within the range of the data, then the effect of As = 1.5 → As = 2.5, is "parametrically" identified.

A statistical model allows us to share information to estimate valid causal effects where we have no data, *provided the model is correct*. This is dangerous and difficult to diagnose, but (caveat emptor) it's possible.



Textbook examples and continuous exposures

The amount of extrapolation and reliance on assuming a correct model grows very quickly as we condition on more and more exposures/confounders, which is part of the **curse of dimensionality**



Continuous exposures beyond textbook examples

The curse of dimensionality is part of the **fundamental problem of mixtures**: correlation between components of a mixture means we should suspect co-pollutant confounding, but adjusting for that confounding in statistical models can lead to problems on its own. To address this information problem, causal inference in mixtures can potentially benefit from:

- Dimension reduction (PCA, clustering, profile regression)
- Shrinkage, selection, Bayesian priors (LASSO, E-net, Bayesian regression/selection, other regularized machine learning)
- Changing the question (WQS, Bayesian kernel machine regression, quantile g-computation)

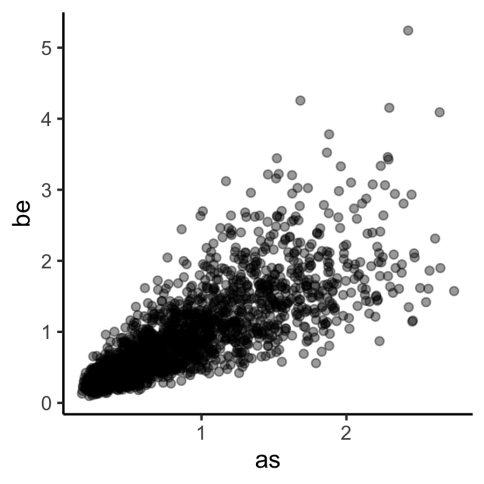
Continuous exposures beyond textbook examples

The curse of dimensionality is part of the **fundamental problem of mixtures**: correlation between components of a mixture means we should suspect co-pollutant confounding, but adjusting for that confounding in statistical models can lead to problems on its own. To address this information problem, causal inference in mixtures can potentially benefit from:

- Dimension reduction (PCA, clustering, profile regression)
- Shrinkage, selection, Bayesian priors (LASSO, E-net, Bayesian regression/selection, other regularized machine learning)
- Changing the question (WQS, Bayesian kernel machine regression, quantile g-computation)

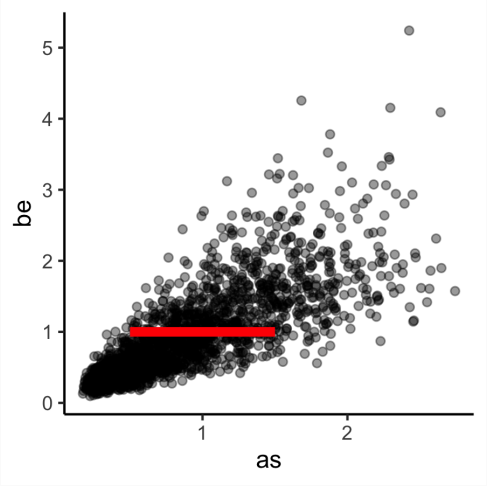
Textbook examples and multiple exposures

Textbook examples of causal inference rarely mention more than one exposure



Textbook examples and multiple exposures

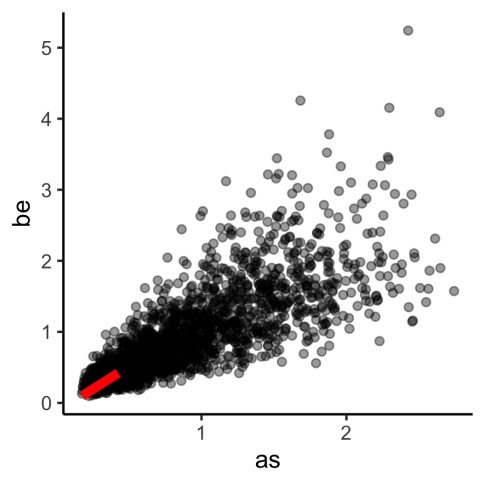
The fundamental problem of mixtures really describes problems with estimating the effect of one exposure, holding others constant:



$$E(y^{a+1}|b=1) - E(y^{a}|b=1)$$

Textbook examples and multiple exposures

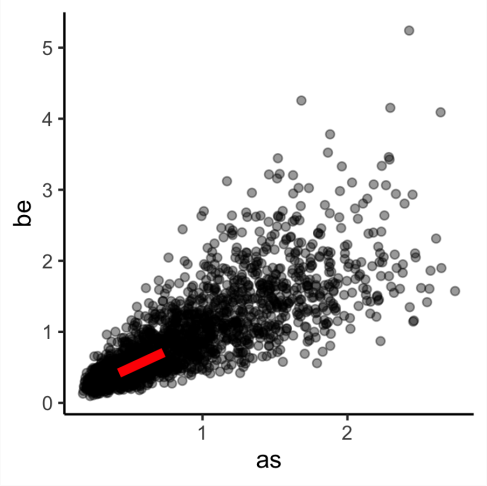
Alternatively, consider the effect of two exposures, where the contrast is again roughly a comparison of the average outcomes at each end of the red line:



$$E(y^{a+1,b+1}) - E(y^{a,b})$$

Textbook examples and multiple exposures

Here we transform exposure so that "1 unit" of exposure is "1 quantile" (e.g. a quartile length)

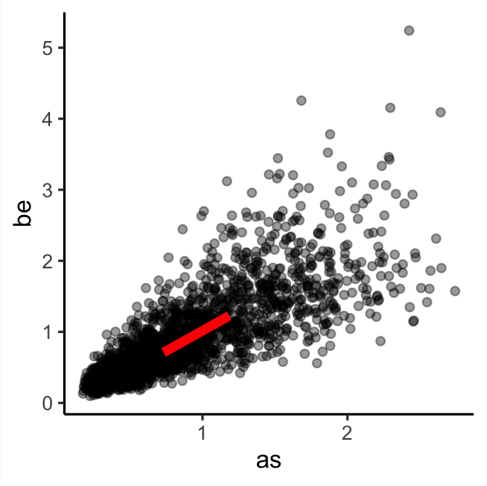


$$E(y^{a+2,b+2}) - E(y^{a+1,b+1})$$

Textbook examples and multiple exposures

This effect is occasionally discussed in the textbook examples as a "joint effect"

see e.g. VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. Epidemiology, 20(6), 863-871.

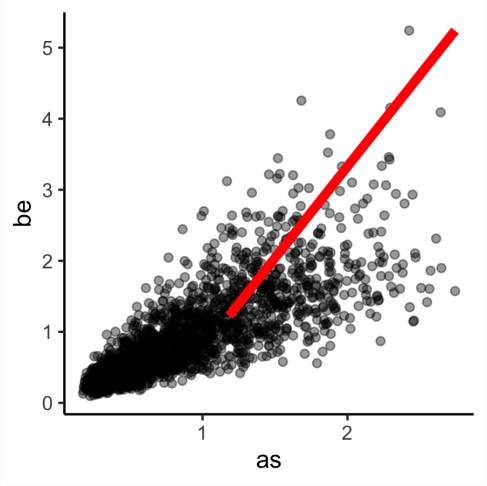


$$E(y^{a+3,b+3}) - E(y^{a+2,b+2})$$

Textbook examples and multiple exposures

This effect is occasionally discussed in the textbook examples as a "joint effect"

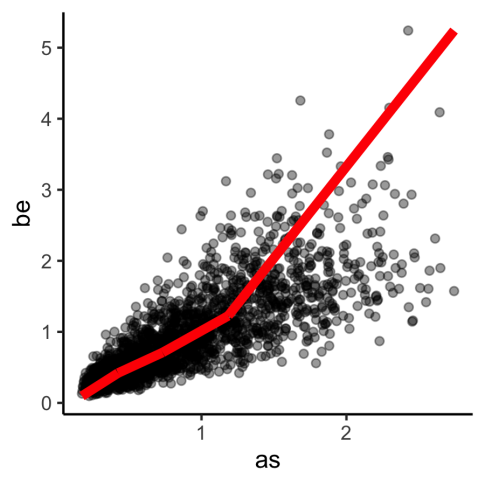
see e.g. VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. Epidemiology, 20(6), 863-871.



$$E(y^{a+4,b+4}) - E(y^{a+3,b+3})$$

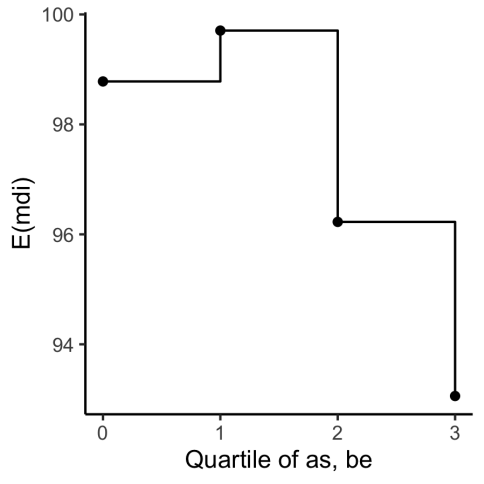
Multiple exposures and marginal structural models

Estimating a series of effects where we make comparisons of the average MDI at the right and left sides of the red line is roughly equivalent to turning our exposures into indicator variables



Multiple exposures and marginal structural models

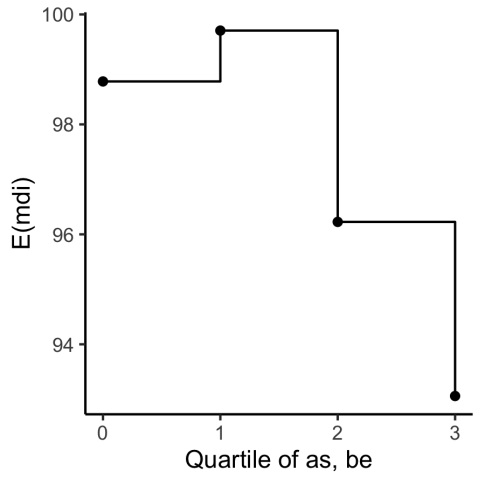
Estimating a series of effects where we make comparisons of the average MDI at the right and left sides of the red line is roughly equivalent to turning our exposures into indicator variables



Multiple exposures and marginal structural models

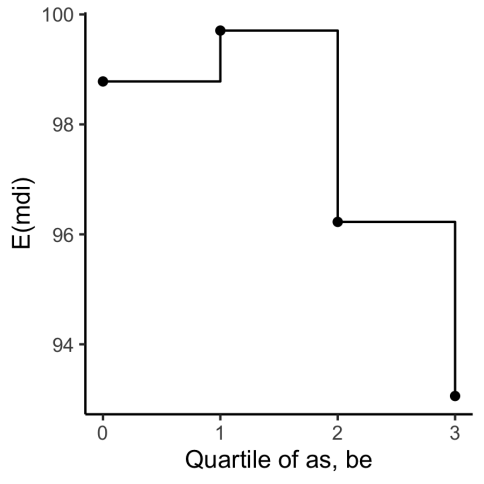
Here we see how the average MDI changes as we increase both arsenic and beryllium by one quantile at a time.

(note that MDI was not on our scatter plot before)



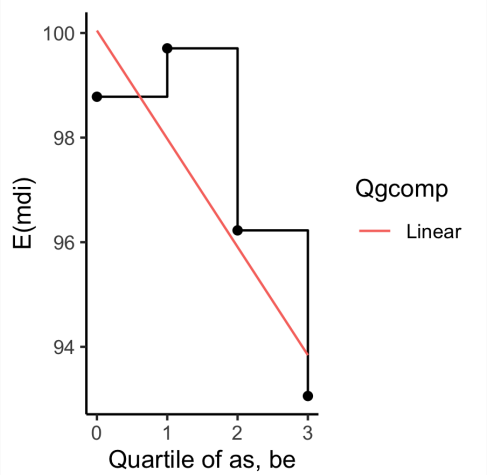
Multiple exposures and marginal structural models

If causal identification conditions are met (substituting our rough "model" for positivity/sparsity), then this can be interpreted as a regression line for a **marginal structural model**, which is just a model for average potential outcomes $E(y^{a,b})$



Multiple exposures and marginal structural models

Methods like Bayesian kernel machine regression[15] and quantile g-computation[16] *can* give estimates from a **marginal structural model**. Quantile g-computation actually fits a model where we assume a parametric marginal structural model, giving a smooth regression line, possibly with a polynomial curve.

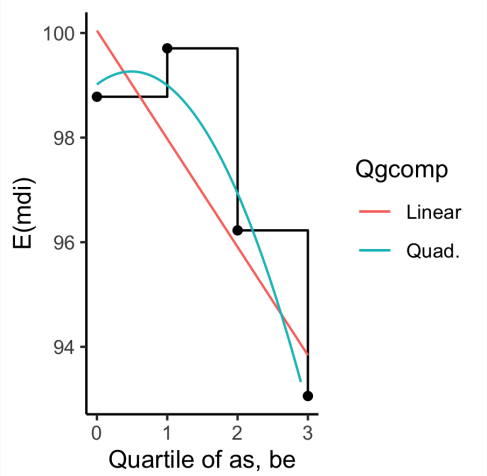


[15] Bobb, Jennifer F., et al. (2015) Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics 16(3), 493-508.

[16] Keil, A. P. et al (2020). A quantile-based g-computation approach to addressing the effects of exposure mixtures. Environmental health perspectives, 128(4), 047004.

Multiple exposures and marginal structural models

Methods like Bayesian kernel machine regression[15] and quantile g-computation[16] *can* give estimates from a **marginal structural model**. Quantile g-computation actually fits a model where we assume a parametric marginal structural model, giving a smooth regression line, possibly with a polynomial curve.



[15] Bobb, Jennifer F., et al. (2015) Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics 16(3), 493-508.

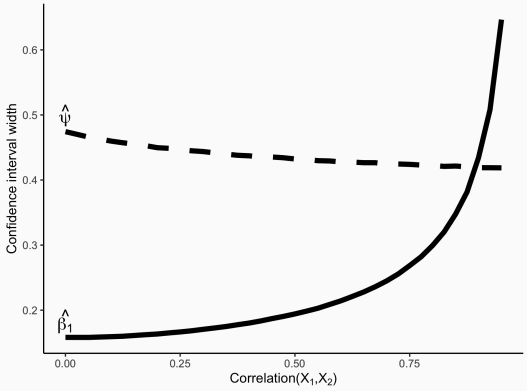
[16] Keil, A. P. et al (2020). A quantile-based g-computation approach to addressing the effects of exposure mixtures. Environmental health perspectives, 128(4), 047004.

Joint effects and the fundamental problem of mixtures

One consequence of the fundamental problem of mixtures is that estimates of independent effects of a given exposure (here β_1) lose precision as correlation ↑

In contrast, marginal structural parameters for the joint effects of multiple exposures (here ψ) can *gain* precision[17]

Estimating joint effects may sometimes be a more efficient use of data than estimating independent effects in search of bad actors



[17] Keil, A. P. et al (2020). A quantile-based g-computation approach to addressing the effects of exposure mixtures. Environmental health perspectives, 128(4), 047004.

Inverse probability weighting and g-computation to estimate joint and independent effects

Inverse probability weighting and marginal structural models

Inverse probability (of exposure) weighting (IPW) is a method for fitting marginal structural models or estimating other "policy" effects such as "how would the average outcome change if we reduced everyone's exposure to zero".

IPW within the mixtures context would rely on having a regression model for each exposure of interest, given confounders and co-exposures, and either an outcome model or a set of individuals who are "representative" of the policy contrast of interest

The basic idea is that the weighted population represents the original study population, had exposures been randomized

Inverse probability weighting and marginal structural models

Methods for weighting with a single continuous exposure are well known, but typically work only in simple examples[18]

For joint effects, one would need a model for each exposure of interest[19]

I cannot recommend use of IPW in the mixtures context for estimating joint or independent effects and do not foresee a way in which current methods could be improved to change this

[18] Naimi, Ashley I., et al. Constructing inverse probability weights for continuous exposures: a comparison of methods. Epidemiology (2014): 292-299.; Kennedy, Edward H., et al. Nonparametric methods for doubly robust estimation of continuous treatment effects. JRSS. Series B, 79(4) (2017): 1229.

[19] Hernán, Miguel A., et al. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. JASA 96.454 (2001): 440-448.

g-computation

G-computation is a generalization of standardization and can be used for fitting marginal structural models or estimating other "policy" effects such as "how would the average outcome change if we reduced everyone's exposure to zero".

For time-fixed exposures, g-computation only requires a standard regression model [20]. Quantile g-computation is a special case.

G-computation extends to time-varying exposures, but requires more modeling[21]

---

[20] Snowden, Jonathan M et al. (2011) Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. AJE 173(7): 731-738.

[21] e.g. Keil, Alexander P., et al. (2014) The parametric G-formula for time-to-event data: towards intuition with a worked example. Epidemiology 25(6): 889.

g-computation

Basic g-computation in the coal plant data:

1. Fit a model for E(mdi | a, b, c, confounders)
2. Make predictions from that model under the "policy" levels of exposure
3. Contrast average predictions for the population or subgroups

Step 1: fit a model

```
mdimod <- glm(mdi ~ as*urbanicity +
  be*urbanicity + cd*urbanicity +
  as*black + be*black + cd*black +
  as*be + as*cd + be*cd, data=coalplant)
```

Step 2: make predictions from that model (e.g. all exposures equal to 1.0)

```
preddata <- coalplant
preddata$as  <- median(coalplant$as)
preddata$be <- median(coalplant$be)
preddata$cd <- median(coalplant$cd)
preddata$pred_med <- predict(mdimod, newdata = preddata)
```

Step 2: (continued)

urbanicity	black	as	be	cd	mdi	pred_med
1	0	0.731	0.709	0.708	82.5	98.1
0	0	0.731	0.709	0.708	93.9	97.2
1	0	0.731	0.709	0.708	123	98.1
0	1	0.731	0.709	0.708	115	97.5
1	0	0.731	0.709	0.708	110	98.1

Here are the first 5 observations with predictions. The predictions represent the expected outcome (MDI) for someone with specific exposure/covariate values

Step 3: Contrast average predictions for the population or subgroups

```
mean(preddata$pred_med)
mean(coalplant$mdi)
```

Taking the mean prediction over the population is, technically, "standardizing" to the empirical distribution of covariates in the data, and yields predicted MDI = 97.7, which is higher than the average observed MDI of 96.9.

The population average MDI is, technically speaking, the expected MDI under the policy "do nothing." So we could contrast that with our expected MDI under the policy "set all exposures to the observed medians" and get an average MDI difference of 0.8, meaning we would expect the population average MDI to increase by 0.8 if we changed everyone's exposures to the median.

Step 3: Contrast average predictions for the population or subgroups

```
mean(filter(preddata, black==1)$pred_med)
mean(filter(preddata, black==1)$mdi)
```

This same contrast among those in the population who self-identified as Black yields an estimate of 1.8. Ignoring uncertainty, this is a larger impact than is observed in the overall population suggesting effect measure modification for the policy effect.

g-computation interpretation

If our causal identification conditions hold, and the linear model we fit is correct, then causal inference tells us that we can interpret our results causally.

This should not be the default - the hard work of causal inference is showing that the identification conditions hold at least approximately.

Without these conditions, g-computation still gives us a standardized effect estimate, which is as valid as any other estimate and has the advantage of directly targeting a scientific question of interest, rather than forcing a question that is answered by regression coefficients.

g-computation: the "underlying fit"

The "underlying fit" for g-computation could be anything, but I (didactically) opted for flexibility by including 9 product terms. (In practice, I use model fit criteria and model dx).

Product terms are difficult to interpret in this context, but note no strong evidence of EMM by race.

	Estimate	Std. Error	t-statistic	p-value
(Intercept)	100.32	1.44	69.46	<2e-16
as	-2.59	2.87	-0.90	0.37
urbanicity	-0.56	1.33	-0.43	0.67
be	-1.49	2.38	-0.63	0.53
cd	-0.90	1.64	-0.55	0.58
black	2.77	1.51	1.83	0.07
as:urbanicity	-3.31	2.36	-1.40	0.16
urbanicity:be	4.61	1.81	2.54	0.01
urbanicity:cd	0.96	1.06	0.91	0.37
as:black	-0.72	2.64	-0.27	0.78
be:black	-2.15	2.09	-1.03	0.30
cd:black	-0.4999	1.2559	-0.398	0.6907
as:be	0.6607	1.3525	0.489	0.6252
as:cd	0.3268	1.1621	0.281	0.7786
be:cd	-0.1557	0.9339	-0.167	0.8676

g-computation: change in average MDI

Using the same underlying model, I estimated more contrasts:

- The independent effects of reducing each exposure to zero (Attributable difference)
- The joint effect of "coal plant decommissioning" by cutting every individual's exposure according to the proportion we would expect based on local coal-plant emissions (96% As, 91% Be, 50% Cd)[20], versus no action (Generalized impact difference)
- Both of the above restricted to Black participants only

[20]These are actually based on estimates for Milwaukee County from the 2014 National Emissions Inventory and used in a forthcoming publication: Keil, A. P., et al. (2021). Bayesian G-Computation to Estimate Impacts of Interventions on Exposure Mixtures: Demonstration with Metals from Coal-Fired Power Plants and Birthweight. AJE

g-computation: extending effect estimates

	Everyone	Black	Non-Black
Attr. mean diff.			
As	2.80 (0.8,4.81)*	3.42 (0.35,6.49)	2.18 (-0.61,4.96)
Be	0.23 (-1.43,1.9)	1.28 (-0.97,3.52)	-0.82 (-3.03,1.38)
Cd	0.58 (-0.5,1.65)	0.94 (-0.62,2.5)	0.21 (-1.22,1.63)
Joint effect			
Shutdown	3.89 (1.81,5.96)	5.91 (2.91,8.90)	1.84 (-0.09,3.76)

Positive values indicate exposure harmful/policy beneficial

*Confidence intervals given in parentheses from bootstrapping (see code)

g-computation: interpretation

Under the causal identification conditions, correct model specification, eliminating arsenic exposure would result in a 2.8 point gain in the population average MDI, an effect which would be stronger among Black participants in the study (3.42).

Provided that we have accurately predicted what would happen to these exposures if local coal-fired power plants were shut down, then Black participants could have experienced a nearly 6 point gain in MDI.

Notably, the underlying model had only weak evidence for effect measure modification by race: these estimates also reflect the fact that Black participants were exposed to higher levels of exposure.

quantile g-computation: a special case

Recall that outcome regression potentially involves a lot of extrapolation for independent effects. Our "joint effect" of a shutdown may similarly involve (to a lesser extent) extrapolation because not all exposures were reduced by similar amounts.

I re-examined the data using quantile g-computation via the qgcomp R-package with a linear underlying model, and stratified by race.

Quantile g-computation estimates a marginal structural model for a 1-quantile (default: quartile) increase in exposure.

quantile g-computation: a special case

	Effect estimate (95% CI)
Everyone	-1.86 (-2.59,-1.14)
Black	-2.71 (-3.69,-1.74)
Non-black	-0.92 (-1.85,0.01)

Note here that a negative value indicates exposure is harmful

Wrapping up

- The hard work of causal inference is in the assumptions: one should take great care
- In our example, it's likely that other pollutants from coal-fired power-plants would affect MDI, so in real examples this simple analysis would be confounded
- In more complex scenarios, we may have modeling problems that require Bayesian solutions to reduce mean-squared error[21]
- Machine learning may similarly address modeling issues[22]
- There are still many technical issues to address inference in high dimensions: this is a fact of our field but should be kept in mind.

[21] This is relatively straightforward: Keil, AP., et al. (2018) A Bayesian approach to the g-formula." SMMR 27(10): 3183-3204.

[22] Oulhote, Youssef, et al. Combining ensemble learning techniques and G-computation to investigate chemical mixtures in environmental epidemiology studies. bioRxiv (2017): 147413.

Wrapping up

However,

- Even if we don't believe the causal identification conditions fully, methods like g-computation are as valid as standard regression approaches
- Methods like g-computation are extremely helpful to take complex modeling scenarios (e.g. a highly flexible model that allows product terms) and still have interpretable results
- Done carefully, focusing on joint effects can reduce some of the statistical/identification issues that arise in mixtures
- G-computation allows us to

Thank you

Causal inference for exposure mixtures

ISEE 2020 virtual pre-conference workshop → ISEE sponsored webinar

---

Alexander Keil
e: akeil@unc.edu
: @PronouncedKeil

January 7, 2021