**Workshop worksheet**

**Hypothetical interventions for exposure mixtures: Practical theory and applications for epidemiologists**

**An ISES/ISEE 2025 pre-conference workshop**

**Barrett Welch, Alex Keil, Maya Spaur**

**The questions here mirror those within comments in the code/output**

**Exploratory data analysis**

1. What are any take-away observations you have from the correlation plots?

*Answer:*

1. How could these correlations influence your approach to
   1. Specifying an exposure mixture (i.e. distinguish between "exposures that are measured together" and "mixture")?

*Answer:*

* 1. Examining the joint effects of changes in the exposure mixture on outcomes?

*Answer:*

1. Modify the code to examine which group is exposed to higher mean levels of arsenic, selenium, barium, uranium, chromium across levels of rural\_urban\_code.
   1. Which group is most highly exposed to each?

*Answer:*

* 1. Do you feel that, overall, exposure to the mixture of these contaminants is equitable across levels of rural\_urban\_code?

*Answer:*

**G-computation:** G-computation has two basic steps: 1) fitting a model (or models) to the data and 2) contrasting predictions from 2 or more hypothetical interventions. The following questions relate to using different types of hypothetical interventions.

**G-computation****example 1: Setting all exposures to the observed median**

1. How do the observed values of biomarker (the outcome) compare to the predicted values under this hypothetical intervention, in which we set the exposures to their median values?

*Answer:*

1. Are there any potential conceptual issues with a hypothetical intervention to set exposures to the median values across all observed data?

*Answer:*

**G-computation****example 2: Proportional reduction of all exposures**

1. What's a technical issue with the proportional reduction approach in which exposures for all individuals are reduced by a constant proportion (Hints: think in terms of causal or modeling assumptions, and consider the post-intervention exposures for the participant with the lowest exposures)?

*Answer:*

1. Regarding the bounded proportional reduction: Which end of the exposure distribution is likely to benefit the most from this type of hypothetical proportional reduction? (i.e. does the intervention potentially benefit those most highly exposed or those who are least exposed?)

*Answer:*

1. Modify the code to re-do our bounded hypothetical intervention, but this time reduce the exposures by 75%. What is the predicted biomarker level under the exposure and the expected percent change for this hypothetical intervention?

*Answer:*

**G-computation****example 3: Capped reduction of all exposures (emulating a regulation or standard)**

1. Was this intervention less or more impactful than what we did above using proportional reductions? What's a potential explanation?

*Answer:*

1. This intervention sets observations EXACTLY to the MCL if they're above. How could this be updated to be a better representation of realistic outcomes/changes?

*Answer:*

**G-computation****example 4: Targeted reduction for exposure equity**

1. After applying the hypothetical intervention (a bounded proportional reduction in exposures): are the geometric mean levels now roughly equal between groups? Why aren't they exactly equal?

*Answer:*

1. What are your take-aways from the equity approach? What factors might we use in the future to better inform the potential impact of real-world interventions to reduce exposure inequities?

*Answer:*

**G-computation****example 5: Bootstrapping to calculate confidence intervals for parametric g-formula estimates**

1. What are the five steps that you repeat to get a bootstrap distribution for effect measures from g-computation?

*Answer:*

1. Examine the point estimates in the pointest object
   1. Which estimates (i.e. intervention and effect measure) are we getting estimates for in this step?

*Answer:*

* 1. How could you interpret the effect measures (point estimates only) in *a* in terms of associations?

*Answer:*

* 1. How could you interpret the effect measures (point estimates only) in *a* in terms of causal effects (if all causal assumptions hold)?

*Answer:*

1. Examine the histogram of the bootstrap sample for the "pct\_change" effect estimate, as well as the two sets of confidence intervals.
2. Does this example show the intervention may have a statistically significant impact on our biomarker outcome?

*Answer:*

1. Does the normality assumption seem reasonable (why/why not)?

*Answer:*

1. Does your answer to *b* match with what you observe from the two sets of confidence intervals?

*Answer:*

1. Which set of confidence intervals would you report?

*Answer:*

1. Do you think your answer to *a* would change if the hypothetical intervention reduced the exposures by 1%? Why or why not?

*Answer:*

1. Which part(s) of this section of code would we change to get bootstrap estimates for the "regulation" intervention (describe only)

*Answer:*

1. **BONUS:** Modify the code to get 90% bootstrap confidence intervals (percentile based). What are they?

*Answer:*

1. **BONUS:** Modify the code to get point estimates and 95% bootstrap confidence intervals for the "regulation" intervention (use the same seed value)? What are they?

*Answer:*

1. **BONUS:** Modify the code to get point estimates and 95% bootstrap confidence intervals for the "regulation" intervention for the binary outcome "cancer" in the data (use the same seed value and a logistic model that has the same terms as the linear model used above). What are they?

*Answer:*