**Group 28 – 1ˢᵗ Project Check-in Discussion**
**Sara O'Brien**

**PRAISE**
- I love this subject area that you all have decided to work in! Environmental risk factors are a large source of health inequities in the Global South, so this is an increasingly pertinent issue for public health professionals to explore.
- I think you all have chosen a credible dataset(s) that will be easy to work with (hopefully making your lives easier!). There is so much data available from this source that I think you will be able to bring in any additional measures you need to consider quite easily.

**ADVICE**
- You state in your SAP that you plan to investigate missing values. Is it possible to supplement these N/A values in the mortality dataset using a previous year, additional source, or other approach? If not, I think it would be valuable to determine why these countries have missing data and if there are any trends in the missingness observed (e.g., many of them appear to be islands, what factors make it more difficult for the World Bank to report these measures in these regions?). I would advise a similar approach for any missingness for the CO2 emissions data. Exploring missingness in this way will help to better define your population (i.e., if you aren't looking at every country but only those with certain characteristics, that is important to acknowledge).
- As someone without much subject matter knowledge on this topic, I am left with a few questions after reading your SAP and project ideas. It is possible some of these questions are unclear to me but have already been considered by your group, but incase you haven't discussed them I think they are important considerations moving forward:
  - When you state that you intend to stratify by regions or groups of countries, will this grouping be based on income (as the secondary question seems to imply) or some other characteristic like geographic region (which your regional comparisons subsection seems to imply)?
  - You state that you intend to consider confounders and effect measure modifiers, but not which covariates you will consider beyond CO2 emissions and income. What additional measures do you plan to include from the World Bank DataBank? How did you/will you select these measures?
  - Are you interested in the main effect of CO2 emissions or the main effect of income on childhood mortality? In some areas it seems like you are more interested in the main effect of CO2 emissions and in other areas it is less clear. I would just be sure to be explicit in what the main relationship is that you are exploring and how other predictors are considered in this relationship.
  - Will you consider CO2 emissions as a mediator in the relationship between income and childhood mortality for your second primary question? In a super quick literature review it seems like there is plenty of evidence of a higher carbon

footprint being downstream of higher income and upstream of higher mortality rates.

- o What type(s) of regression model(s) are you exploring? The measures you're looking at make a linear model seem intuitive, but is there a linear relationship in your data based on your EDA? We haven't covered this in the course yet, but it would also be fun to do survival analyses with this data.

- It could be interesting to explore how associations between CO2 emissions, income, and mortality have changed over time, given the increased burden of climate change on many communities in the past few decades. I think it would be especially compelling to show that an association exists and continues to be exacerbated by a lack of policy intervention, lack of emphasis in development goals, etc. to reduce emissions.

**CONCERNS**

- I think this project has a lot of potential and will turn out great. My only concerns can be addressed by ensuring that the questions that I (+ the teaching team and other reviewers) posed are addressed :)