

### **PRAISE**

- I like the general idea of your statistical analysis. Associations to CO<sub>2</sub> emissions is certainly relevant to current societal issues, and I've not seen a primary analysis conducted with an outcome of infant-specific mortality. I'll be very interested to see your results!
- This is a cool use of this database, as it is so extensive. I like that you've focused on a specific question but are utilizing a number of covariates provided from other databases. I'm very curious to see how you plan to construct a dataset in combination of covariates from several databases and what selection criteria you'll use for the covariates.
- I think that seeking guidance from the Boston Children's Hospital Global Health research team will be a great resource for your group. With similar research likely to have been conducted by them, I believe they'll be very helpful to utilize throughout the project.

### **ADVICE**

- Are you specifically looking at the gender statistics, as the link in your discussion board post follows to? You don't mention anything about gender comparison in your primary or secondary analyses, so I was wondering if that will take a role as a covariate at all.
  - I cannot find data regarding CO<sub>2</sub> emissions in the gender-specific data, but I can when looking at the SDGs database (is this what you're utilizing?). Further, there are 4 variants of defining CO<sub>2</sub> emissions, which type are you using? What is your operational definition of this predictor?
  - In the SDGs database, there are also numerous definitions of an income variable. Which of those definitions are you utilizing? What is defined as "high" or "low" income, or are you reclassifying based on the given data? (e.g., average individual net income, average household income, categorization established by the dataset, etc.).
- Be sure to be as specific as possible in defining your exposures/covariates of interest (CO<sub>2</sub> emissions and "higher" income for your primary analyses) and outcome (in what unit is the mortality rate defined, beyond the age, and how will the deaths be stratified) to avoid confusion from individuals reading through your analyses and to limit bias in interpretation of your variables.
- When you state that you want to conduct stratified analysis based on region/groups of countries, what are you referring to? Is this simply classifying by individual country or grouping nations together based on external criteria or geographical area? Again, be sure to operationally define these prior to conducting your analysis.
- Are you looking at a specific time period (e.g., the year 2020) or all the time frames provided in the dataset (all years provided by the database)? It may be of worth to

examine results stratified by time-period (in addition to by country) as another secondary question since the statistics of particular countries may have changed based on policy change/regulations (e.g., countries part of the UN may see large change throughout the data collection with the addition of SDGs).

- For your regression analysis section, there's no specification of what kind of regression you'll be utilizing (e.g., linear, logistic, survival analysis, implementing splines, GAM, etc.). With continuous child mortality as your outcome, it could fall into linear or survival analysis (depending on how you ultimately define child mortality), will you be conducting multiple types of regression? It may be worth fitting and comparing several types of regression models if your classifications of variables can fit in such a manner.
  - In your model selection section, you indicate that you'll evaluate goodness of fit via adjusted  $R^2$  and other characteristics. However, it's important to consider that, depending on the regression technique you decide to utilize, adjusted  $R^2$  is a characteristic provided only in select models, so you may want to consider examining a more standardized criterion.
  - Similarly, when checking model assumptions, note that the assumptions may differ depending on the type of regression you decide to utilize.

## **CONCERNS**

- You indicated that there are 180 observations (i.e., 180 countries being examined), whereas I can see 261 countries in this dataset, is there a particular reason for the exclusion of 81 countries?
- You claim to have about 25 covariates under consideration, but this may be too many to include in your final modeling, as there should minimally be about 10 observations per covariate of interest. The use of this many covariates with only examining 180 countries may lead to an introduction of bias in your analyses.
- When looking at each classification of CO<sub>2</sub> emissions in the databank you provided, the values are voided for every country. Be sure that whichever database you do select has minimal missing values of each of your covariates, this is especially concerning for your primary covariate of interest.