Tentative analysis plan:

1. Identify main dependent variables – cardiovascular disease outcomes:
   1. Definitely include
      1. Hypertension (1471 cases)
      2. Heart attack (89 cases)
   2. Possibilities
      1. Non-specific heart problems (214 cases)
         * Weaker evidence after adjusting for both childhood and adult SES
      2. Congestive heart failure (56 cases)
         * Weaker evidence after adjusting for both childhood and adult SES
      3. Stroke (84 cases)
         * Weaker evidence after adjusting for both childhood and adult SES
2. Analyses
   1. Survival analyses may be most informative, since all outcomes have date of event or diagnosis (except non-specific heart problems variables)
   2. Split sample by sex, compare estimates of confidence intervals of variables, across models.
      1. This is my preference on how to do these sorts of analyses as I often find using sex as an interaction variable makes the models more difficult to interpret.
   3. With this many variables, my inclination would be to run these through a machine learning process like the LASSO or Gradient Boosting, separately for each sex, then confirm the models in a more traditional survival analysis
      1. Possibly combine the sexes at this point and introduce interactions as informed by the machine learning models
3. Control variables
   1. Childhood age
   2. Childhood SES
   3. Adult SES
   4. Education?
4. Variables to test sensitivity
   1. Health care use
      1. Does R have health care provider to see when needed
      2. Time since last physical exam
      3. Specific medical tests
         * Blood pressure
         * Cholesterol
         * Blood sugar/diabetes
   2. Weight, height -> BMI
   3. Diet?
   4. Exercise?
   5. Alcohol use
   6. Smoking
   7. Drug use
   8. Sleep