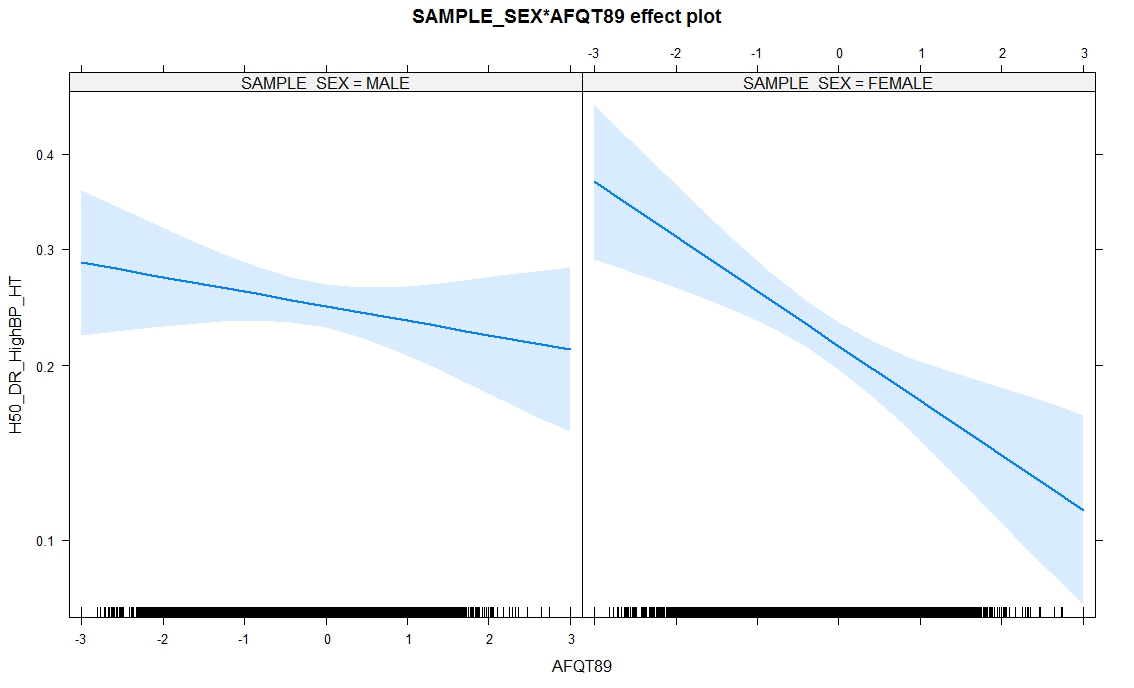
Analysis outline:

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. Preliminary interaction plot



X-axis is an individuals’ score on the Armed Forces Qualification Test, taken between the age of 15 and 23 years. Y-axis is the probability that an individual will have been diagnosed with hypertension by age 50. This includes controls for Child and Adult SES.

1. Possible formal 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.
   3. With so many possible control variables (see below), my inclination would be to run these through a machine learning process like the LASSO or Gradient Boosting, then confirm the models in a more traditional survival analysis.
      1. Sex interactions can be informed by the machine learning models as well.
2. Control variables
   1. Childhood age
   2. Childhood SES
   3. Adult SES
   4. Education
3. 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
4. Open Questions
   1. Whether to split models, or focus on identifying sex interactions within models of the entire sample?
   2. How far to go in the search for other variables that can explain the sex differences?
   3. How to present different outcome variables? Model them all in the same fashion and assess for consistency?
   4. Others?