**Introduction**

(work in progress)

**Methods**

**Dataset Description**

The Baby's First Years (BFY) project is first randomized controlled trial (RCT) in the U.S. designed to evaluate the causal impact of poverty reduction on a child's early development. Since its initiation in 2018, the BFY has recruited 1,000 mothers of infants with incomes below the federal poverty line across four diverse communities: New York City, New Orleans, the greater Omaha metropolitan area, and the Twin Cities. Mothers were recruited from postpartum wards shortly after giving birth and received a monthly cash gift by debit card for the first 76 months of their child's life. Mothers were randomly assigned to one of two groups: an experimental group (n = 400) receiving $333 per month ($3,996 per year) and a control group (n = 600) receiving $20 per month ($240 per year). Importantly, participants did not lose eligibility to public benefits (e.g. Supplemental Nutrition Assistance Program, Head Start, or Medicaid) due to the cash reward (**Nobel et al. (2021)**).

The inclusionary criteria was the following: (1) mother’s self-reported income was below the federal poverty threshold in the previous calendar year; (2) mother was of legal age for informed consent; (3) infant was admitted to the newborn nursery and not requiring admittance to the intensive care unit; (4) mother was residing in the state of recruitment; (5) mother reported not being "highly likely" to move to a different state or country in the next 12 months; (6) infant was discharged in the custody of the mother; and (7) mother was either English or Spanish speaking (necessary for instruments of some child outcomes) (**Nobel et al. (2021)**).

Families in the BFY study were involved in four waves of data collection. First, baseline data was collected in the hospital shortly after birth. Afterwards, in-person home visits were conducted when the child was 12 and 24 months of age. Lastly, a university-based laboratory visit was conducted when the child was 36 months of age. This analysis used self-reported surveys data collected at baseline including mother demographics, mother-father relationship, and public assistance as predictors of language outcome (See Table 1). The language outcome of interest was the communication subtest of the Ages and Stages Questionnaire (ASQ) collected at 12 months of age. The ASQ is a developmental screening tool designed to assess young children's progress across five key domains: Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-Social. The Communication domain specifically evaluates a child's ability to understand and use of both expressive and receptive language (**Nobel et al. (2021)**).

**Dataset Access and Cleaning**

Variables in the baseline data file are of two types – raw and generated. The first type of variables is considered raw because they are direct outputs from the Baseline and Screening survey program from SRC. They are unprocessed.

The second – “generated” – type of variables in the Baseline\_Clean\_Data\_BFY data file are generated by BFY analysts in preparation for analyses of the data. These variables are re-coded (e.g., yes/no responses are coded yes=1 and no=0). In addition to simple recoding of values, a number of quality checks were conducted to create complicated generated variables, such as income, that required analytic decisions.

**Exploratory Analyses**

**Univariate Analyses**

****

**Bivariate Analyses**

**Model Development**

**Model Selection**

**Hyperparameter Tuning**

**Model Performance**

**Evaluation on Testing Set**