Geographic and Social Predictors of Readmission.

Author One 1, …, …, …, Pseudonym Elemental PhD, MD 1,✉

2021-11-17

[Provide a summary of objectives, study design, sample size, predictors, outcome, statistical analysis, results, and conclusions.]

# Preliminary Results

## Data

As proof of principle, we used AHRQ-curated census data from the years 2015 and 2016, which we will extend to all available years during the proposed study. Using zipcode tabulation areas (ZCTAs) we crosswalked the AHRQ data to the Social Deprivation Index [reference] and to our own calculations of percentages of hospital discharges that did not result in readmissions.

## Analysis

Before viewing the merged dataset in any way, we randomly assigned each ZCTA to a training (N=???) or testing (N=???) subset (each with 50% likelihood). Variable selection and all other decisions about data analysis were made using only the training subset, having blinded ourselves to the testing subset until the entire analysis pipeline was finalized.

Our previous research suggests that the following socieconomic variables are possible predictors of readmission (the variable named NoReadmission ) : ACS\_PER\_CAPITA\_INCOME, ACS\_PCT\_POSTHS\_ED, ACS\_MEDIAN\_HOME\_VALUE, ACS\_PCT\_MEDICAID\_ANY, ACS\_PCT\_DISABLE, and sdi\_score. However, we had no reason to believe these are the only relevant variables nor that their effect is additive (i.e. we cannot exclude the possibility that the effect of a variable might increase or decrease depending on the values of other variables).

We used bi-directional stepwise selection with the Bayes Information Criterion (BIC) to select the terms for a regression model. The models considered ranges from no predictors to all available variables with all possible two-way interactions. The model selected through this process was as follows:

'[model summary goes here]'

'[residual plots go here]'

'[plot of predicted vs observed values goes here]'

We then made predictions on out-of-sample data (i.e. the training set) without re-fitting or altering the original model and got the following result:

'[plot of predicted vs observed training values goes here]'