The goal of this project is to determine how the local uninsured rate influences the readmission rate of hospitals and whether good patient counseling, as judged by the patient, has any effect.

The audience for this analysis is both hospital administrators and those that have an interest in insurance policy. If an administrator finds the local population to have a high uninsured rate, the ability to mitigate this with additional efforts targeted toward discharge counseling could be a viable technique to lower readmission rates. The readmission rate is a key indicator of quality of care and healthcare facilities are penalized by Medicare for high readmission rates. These penalties result in reimbursements being withheld and in October of 2016 these penalties were increased by 20 percent.

To increase the quality of patient outcomes, advocates for increased insurance coverage may be interested in the analysis as evidence that increases in insurance coverage lower readmission rates. Lower readmission rates have the effect of lowering overall healthcare costs.

The data used for this analysis will be the Small Area Health Insurance Estimates by the US Census Bureau (https://www.census.gov/did/www/sahie/data/20082014/index.html) and the Hospital-level HCAHPS data and readmission rate date available from Medicare (https://data.medicare.gov/data/hospital-compare).

The first step is using the SAHIE data to determine the mean uninsured rate for each county in the United States that also has a hospital. Once the mean uninsured rate of the population served in each county it can then be correlated to the mean readmission rate for hospitals in that county. If there is a correlation, then the results of the HCAHPS survey can be used to determine if there is a correlation between discharge counseling and readmission rate among hospitals that have similar rates of uninsured patients.

The key pieces of data in the Medicare data sets are the state and county, the readmission rate for each facility, and the HCAHPS scores for the metric associated with patient education. In the SAHIE data the key data fields are the state and county as well as the estimated uninsured rate in each county.

One of the limits of the data SAHIE data is that is the uninsured rate is an estimate. Ideally, this data would be a more exact metric, but SAHIE estimates are the accepted as a measure of the true uninsured rate. Another limitation of the data is the three data sets do not measure the same period. The SAHIE data and the Medicare HCAHPS data follow calendar years, January – December, whereas the period measured by the Medicare readmission data is July – June. This issue is overcome by weighting each year of the SAHIE data and HCAHPS data 50% and using the resulting metric to compare to the Medicare data, as half of the Medicare data falls into each of a pair of consecutive years. Another issue with the data is the granularity of the uninsured rate. The uninsured rate is aggregated at the county level while the data for readmission and HCAHPS scores are collected at the hospital level. Any correlation would be more meaningful if all data was at the hospital level, avoiding the need to aggregate hospital-level metrics at the county level and yielding less accurate results. Additionally, potential user of the analysis would be handicapped by the aggregation as many of the changes that could affect the readmission rate would take place at the hospital level. A missing element of the data set that would be useful would be bins for age ranges. As it is there is no way to take age into consideration, conclusions can’t be drawn regarding the impacts on various age groups.

The first step in munging the data is combining the years being analyzed into one superset for each data set. Then it is necessary to remove the extraneous header rows from each set. In the Medicare data set that includes readmission rate, the columns that are only meaningful at the hospital level can be removed. For example, the street address and the provider number are unnecessary as the data will be analyzed at the county level. The rows to be removed are those that have data related to condition-specific readmission, such as CHF or pneumonia. This analysis deals only with hospital-wide readmission rate. The second Medicare data set, which includes the HCAHPS scores for each facility, can likewise have the data removed that deals only with hospital level data points. Additionally, the rows with those metrics that score data other than patient education can be removed. The SAFIE data set includes many population categories, such as race, sex, and income, but because the Medicare data does not, all fields that designate specific populations can be removed. As a result, the columns that estimate uninsured rates for the various populations can also be eliminated.

The first Medicare data set needs to be grouped by the county/state combination and the time over which the data was collected. An average then needs to be taken from the all readmission rates within each county over each year. The second Medicare data set needs to be grouped the same way and an average taken of the relevant patient education performance measures for each county/state combination. The SAHIE data set needs all population-specific categories removed leaving only the overall uninsured rate estimates for the population of each county.