Effects of HCAHPS Scores and Area Uninsured Rate on Hospital Readmission Rate

Tim Harvey

05.31.2017

Objective

- The goal of this project is to determine how the local uninsured rate influences the readmission rate of hospitals and whether an increase in any of the HCAHPS linear scores has an effect.

Audience

- 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 the most predictive HCAHPS scores to lower readmission rates.
- 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 decreasing overall healthcare costs.

Data and Data Issues

- The data used for this analysis will be the Small Area Health Insurance Estimates (SAHIE) by the US Census Bureau (https://www.census.gov/did/www/sahie/data/20082014/index.html), and the hospital-level HCAHPS and readmission data available from Medicare (https://data.medicare.gov/data/hospital-compare).
- The two main issues with these data sets were the misalignment of the time periods and the aggregation to the county level rather than being able to have hospital-level data for all three data set.
- The first issue was addressed by using a weighted average of consecutive years to align the time periods. The second issue was addressed by aggregating the hospital-level data to the county level. IN the end this as inconsequential as the SAHIE data had a very low correlation to readmission rate and was dropped, allowing for analysis to occur at the hospital level.

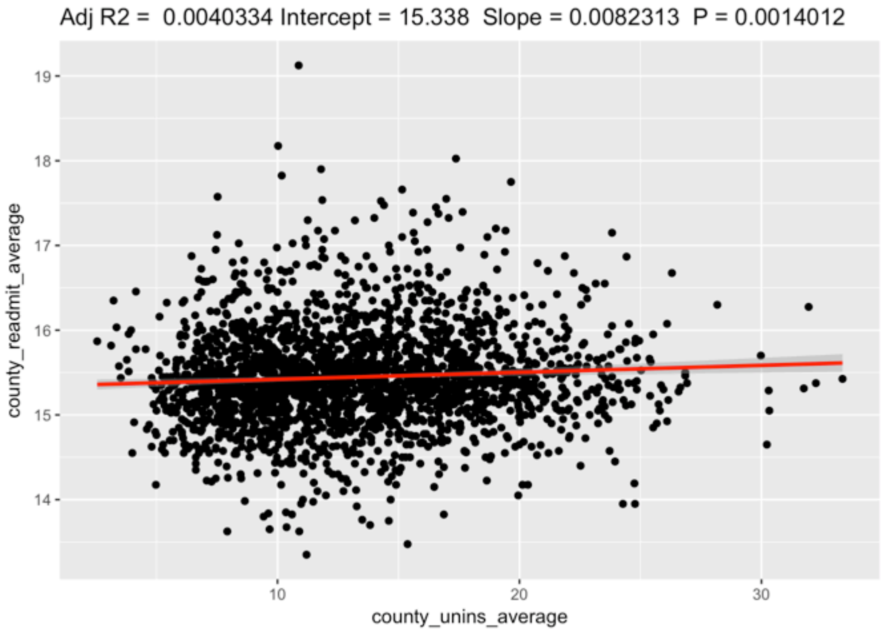
Munging the Data

- The vast majority of SAHIE data was for subsets of the population. For example, the data was divided by age, sex, race, and income. These were discarded in favor of the general population metrics. The county and state name required for joining the other data set was in a different format and had to be reformatted to all uppercase and removing designations such as “County” and “Parish”. Consecutive years were weighed using 0.75 and 0.25 respectively to align the measure with the HCAHPS data.
- The HCAHPS data was formatted using the spread() function in dplyr to transform the measure name into variables that could be evaluated independently.
- Hospital-wide readmission data was extracted from the Medicare readmission data by filtering the READM_30_HOSP_WIDE measure. Consecutive years were weighed using 0.25 and 0.75 respectively to align the measure with the HCAHPS data.

Initial Analysis

- The summary and scatterplot to the left show correlation with only the county uninsured rate considered. The adjusted R squared is less than 1% and was discarded in favor of making a more granular hospital-level analysis.

```
##
## Call:
## lm(formula = county_readmit_average ~ county_unins_average, data = analyze_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0802 -0.3497 -0.0505  0.3087  3.6974
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         15.338042   0.036441 420.896   <2e-16 ***
## county_unins_average  0.008231   0.002574   3.198   0.0014 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5939 on 2278 degrees of freedom
## Multiple R-squared:  0.00447,	Adjusted R-squared:  0.004033
## F-statistic: 10.23 on 1 and 2278 DF,  p-value: 0.001401
```



Initial Analysis Continued

- When the HCAHPS data was aggregated to the county level the highest adjusted R squared value for any single attribute was 0.08
- When the data was summarized at the hospital level the adjusted R squared values were significantly higher.

```
[1] "H_CLEAN_LINEAR_SCORE"
[1] 0.02409226
[1] "H_COMP_1_LINEAR_SCORE"
[1] 0.03833288
[1] "H_COMP_2_LINEAR_SCORE"
[1] 0.01506094
[1] "H_COMP_3_LINEAR_SCORE"
[1] 0.062421
[1] "H_COMP_4_LINEAR_SCORE"
[1] 0.02963111
[1] "H_COMP_5_LINEAR_SCORE"
[1] 0.03796599
[1] "H_COMP_6_LINEAR_SCORE"
[1] 0.0817606
[1] "H_COMP_7_LINEAR_SCORE"
[1] 0.06167917
[1] "H_HSP_RATING_LINEAR_SCORE'
[1] 0.05313558
[1] "H_QUIET_LINEAR_SCORE"
[1] 0.006334165
[1] "H_RECMND_LINEAR_SCORE"
[1] 0.05960038
[1] "county_unins_average"
[1] 0.004033388
```

```
[1] "H_CLEAN_LINEAR_SCORE"
[1] 0.05954117
[1] "H_COMP_1_LINEAR_SCORE"
[1] 0.08409356
[1] "H_COMP_2_LINEAR_SCORE"
[1] 0.05855484
[1] "H_COMP_3_LINEAR_SCORE"
[1] 0.1062389
[1] "H_COMP_4_LINEAR_SCORE"
[1] 0.08251743
[1] "H_COMP_5_LINEAR_SCORE"
[1] 0.07506323
[1] "H_COMP_6_LINEAR_SCORE"
[1] 0.1160424
[1] "H_COMP_7_LINEAR_SCORE"
[1] 0.1115123
[1] "H_HSP_RATING_LINEAR_SCORE"
[1] 0.1074239
[1] "H_QUIET_LINEAR_SCORE"
[1] 0.03888186
[1] "H_RECMND_LINEAR_SCORE"
[1] 0.1047647
```

Creating the Model

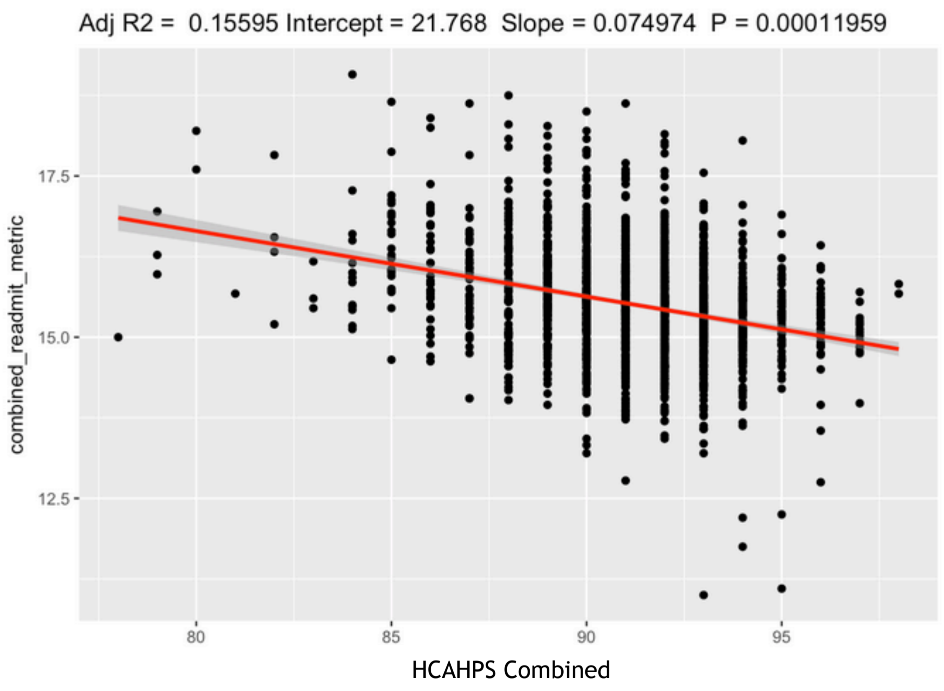
- The adjusted R squared for the regression with all variables is .1586. By eliminating 5 of the variables the adjusted R squared only drops to .1553

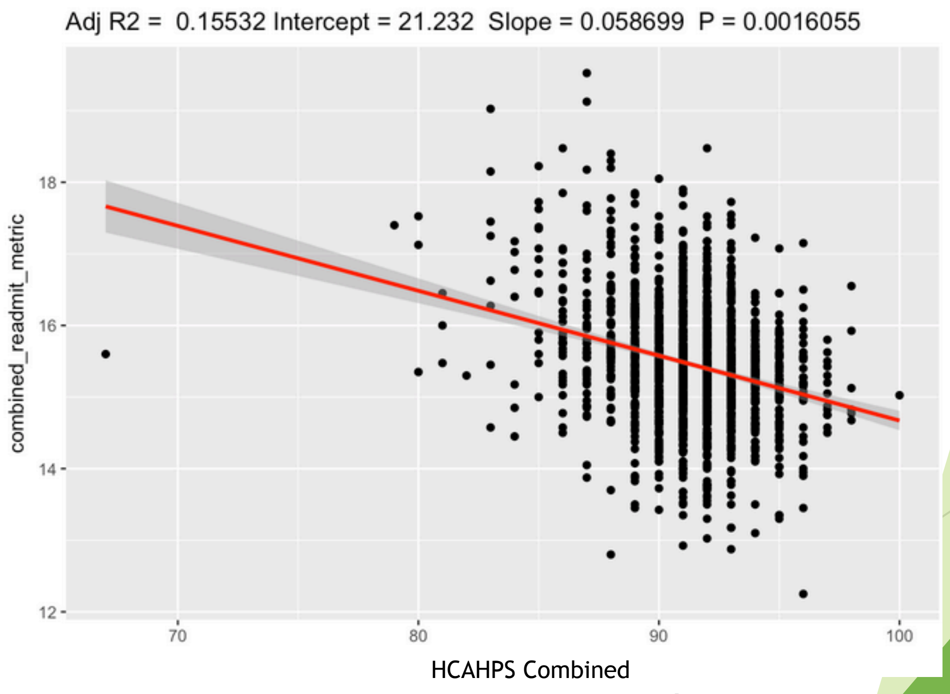
```
##
## Call:
## lm(formula = combined_readmit_metric ~ ., data = data_final_train[c(2:13)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7712 -0.4979 -0.0195  0.4733  3.2149
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              21.370789   1.038072  20.587  < 2e-16 ***
## H_CLEAN_LINEAR_SCORE     -0.013295   0.007274  -1.828 0.067764 .
## H_COMP_1_LINEAR_SCORE     0.073207   0.020516   3.568 0.000369 ***
## H_COMP_2_LINEAR_SCORE    -0.002242   0.012923  -0.173 0.862302
## H_COMP_3_LINEAR_SCORE    -0.048280   0.009579  -5.040 5.14e-07 ***
## H_COMP_4_LINEAR_SCORE    -0.021253   0.014919  -1.425 0.154459
## H_COMP_5_LINEAR_SCORE     0.026685   0.008383   3.183 0.001482 **
## H_COMP_6_LINEAR_SCORE    -0.058019   0.007268  -7.983 2.59e-15 ***
## H_COMP_7_LINEAR_SCORE     0.010294   0.016502   0.624 0.532853
## H_HSP_RATING_LINEAR_SCORE 0.033441   0.020765   1.610 0.107481
## H_QUIET_LINEAR_SCORE     -0.011814   0.005258  -2.247 0.024782 *
## H_RECMND_LINEAR_SCORE    -0.058376   0.013499  -4.325 1.62e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7719 on 1724 degrees of freedom
##   (427 observations deleted due to missingness)
## Multiple R-squared:  0.1639, Adjusted R-squared:  0.1586
## F-statistic: 30.73 on 11 and 1724 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = combined_readmit_metric ~ ., data = data_final_train[c(2,
##     4, 6, 8, 9, 10, 13)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7699 -0.5123 -0.0180  0.4628  3.3521
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            2.123e+01  8.667e-01  24.497  < 2e-16 ***
## H_COMP_1_LINEAR_SCORE  5.870e-02  1.858e-02   3.160  0.00161 **
## H_COMP_3_LINEAR_SCORE -5.468e-02  9.015e-03  -6.065 1.62e-09 ***
## H_COMP_5_LINEAR_SCORE  2.256e-02  8.128e-03   2.776  0.00556 **
## H_COMP_6_LINEAR_SCORE -5.234e-02  7.068e-03  -7.405 2.03e-13 ***
## H_COMP_7_LINEAR_SCORE -9.806e-05  1.605e-02  -0.006  0.99512
## H_RECMND_LINEAR_SCORE -4.197e-02  8.053e-03  -5.212 2.09e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7734 on 1729 degrees of freedom
##   (427 observations deleted due to missingness)
## Multiple R-squared:  0.1582, Adjusted R-squared:  0.1553
## F-statistic: 54.17 on 6 and 1729 DF,  p-value: < 2.2e-16
```

Testing the Model

- The training and test sets were create by randomly assigning half of the data to each set. The adjusted R squared deferred between the two sets by just .00063.





Conclusions

- Clearly there is not a strong correlation between any one of the HCAHPS factors and readmission rate and only a slight correlation using the best combination of factors.
- That said, the analysis can still of use to those aiming to reduce readmission rate.
- With limited resources focus is best applied to improve those metrics that measure how well patients feel they are educated upon discharge (H_COMP_6_LINEAR_SCORE, H_COMP_7_LINEAR_SCORE) and the responsiveness of the staff (H_COMP_3_LINEAR_SCORE).
- While the analysis of the HCAHPS data is not necessarily predictive, it does offer guidance as to which secondary factors focus may be best applied to improve patient outcomes.