

Predicting MRSA Healthcare Related Infections

Jeffrey Jex

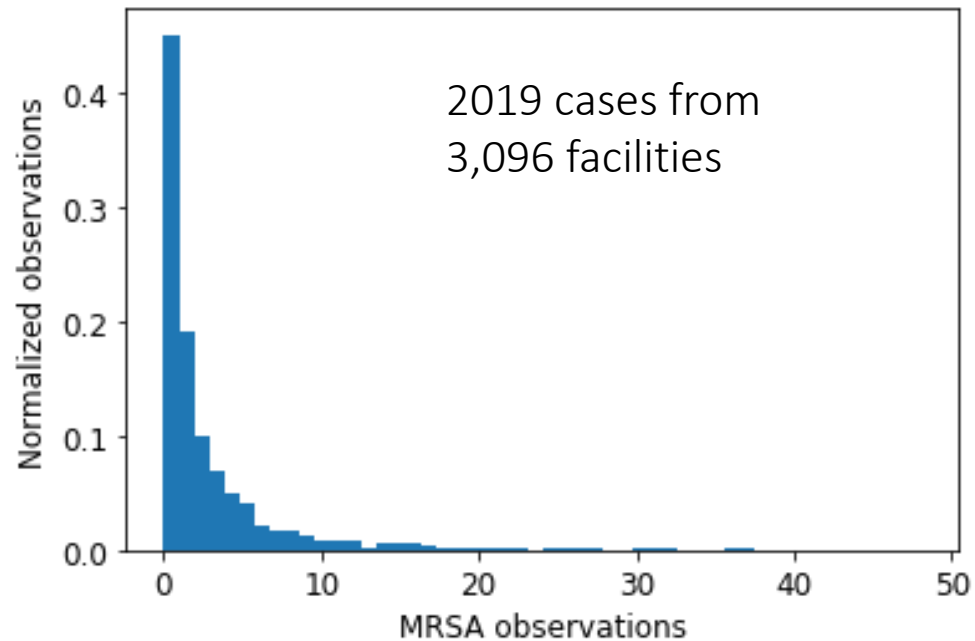
Capstone project #1

May 2021

"At any given time, about 1 in 25 inpatients have an infection related to hospital care. These infections lead to tens of thousands of deaths and cost the U.S. health care system billions of dollars each year."

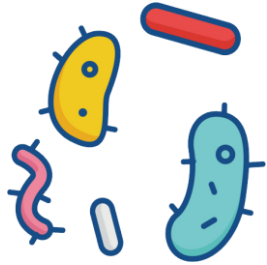
Goal

- Predict 2019 MRSA infections at the hospital level, using only 2019 data



Data Sources

Joined data from 3 separate CMS (Center for Medicaid and Medicare Services) sets



HAI data



Patient Surveys



Timely and effective
care

Data Cleaning

CMS data

Healthcare Associated Infections (HAIs)

- 4878 facilities down to 1715 facilities

Patient Surveys (HCAHPS)

- Remove partial year surveys and star ratings
- Save only aggregated question results
- From 93 features down to 11

Timely and Effective Care

- Only include sepsis care
- From 16 features down to 5

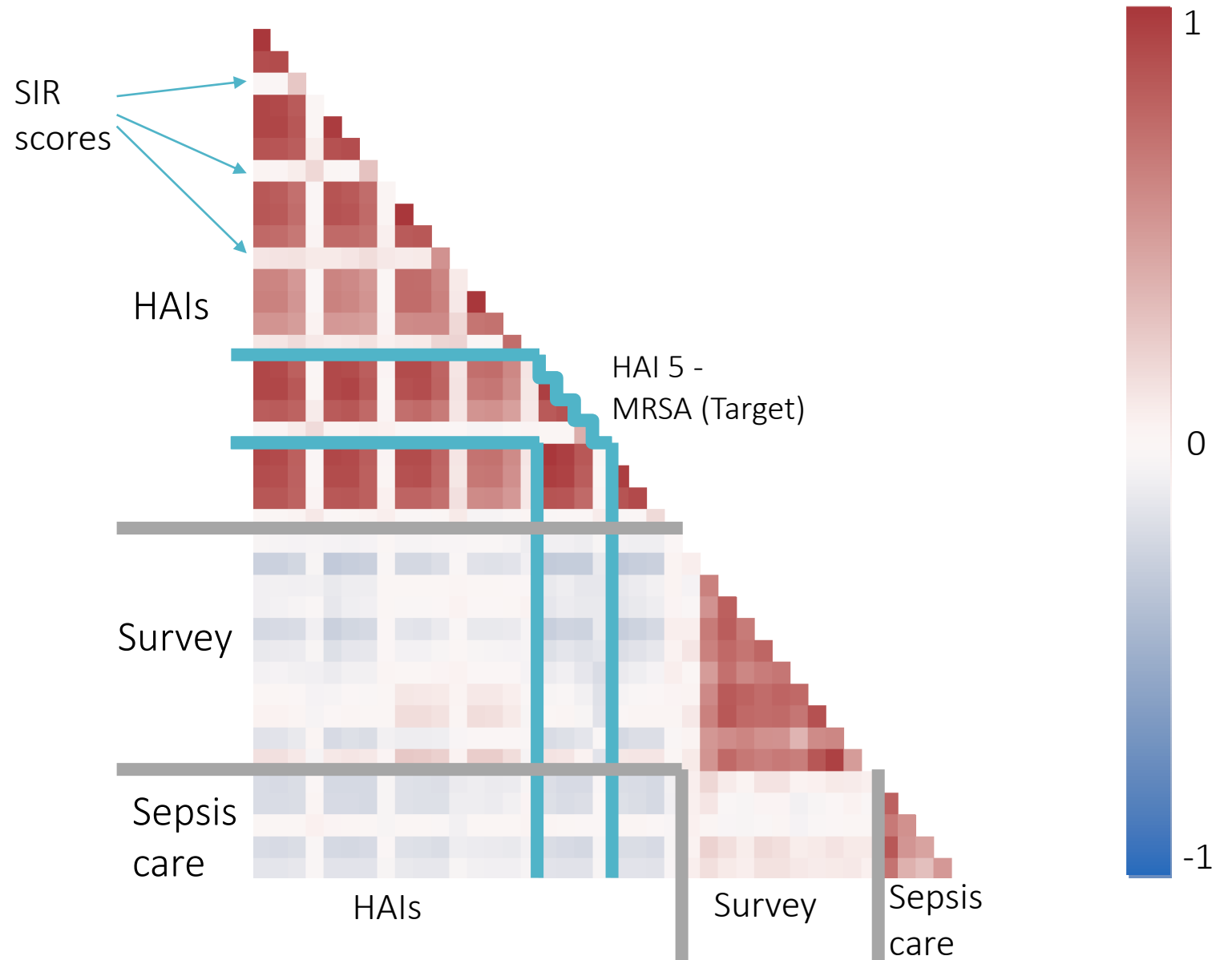


EDA

Correlation heat map

- strongest correlations are internal to each dataset

$$\text{Standardized Infection Ratio (SIR)} = \frac{\text{Actual Number of Infections}}{\text{Predicted Number of Infections}}$$



Modeling

Avoid data leakage – removed following features from predictors:

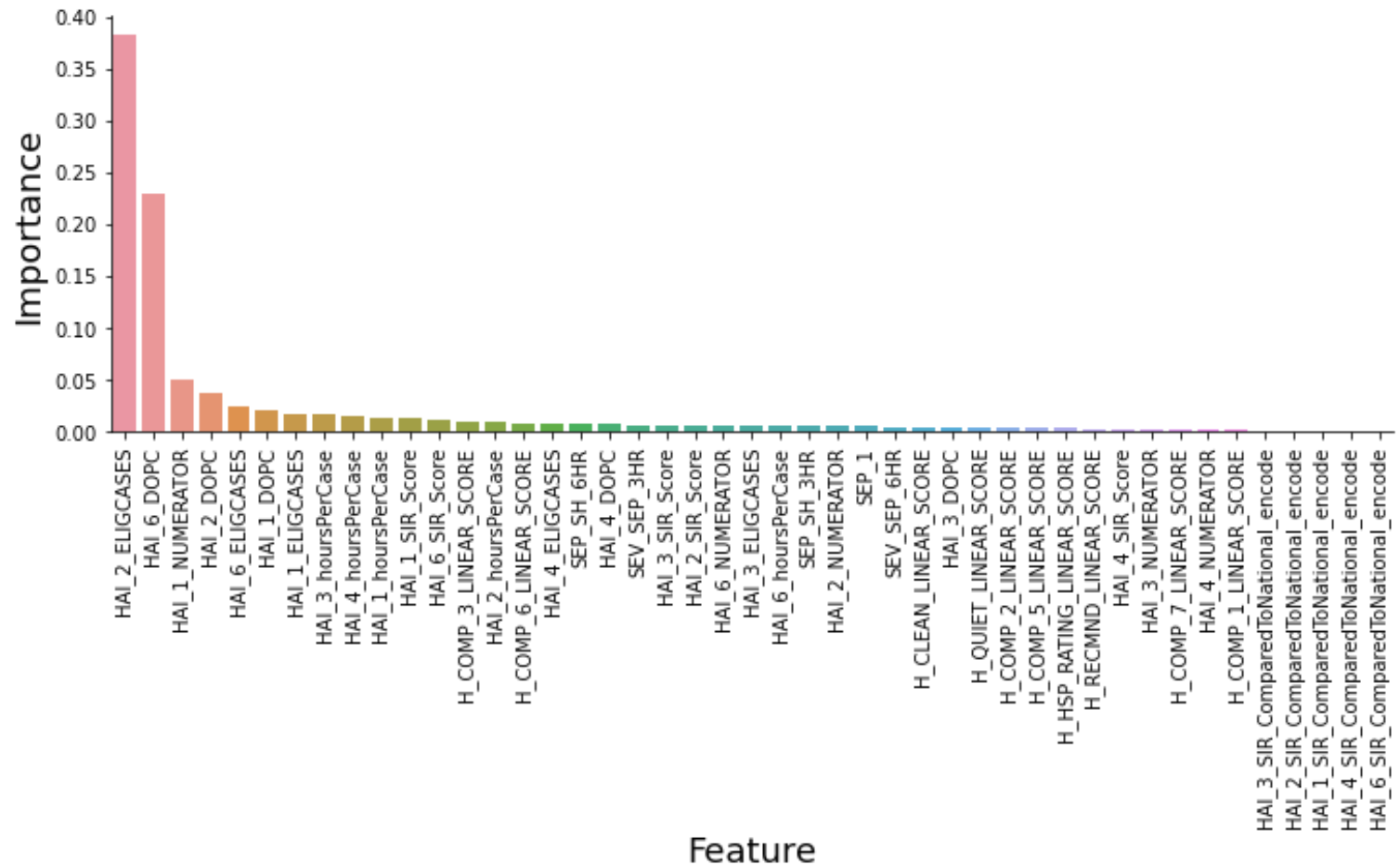
- **MRSA observed cases – target variable**
- MRSA predicted cases (by the CMS)
- MRSA SIR score
- MRSA total patient hours
- MRSA SIR compared to national category

Algorithms tried – OLS regression, ridge regression, decision tree, random forest

Random Forest- feature importance

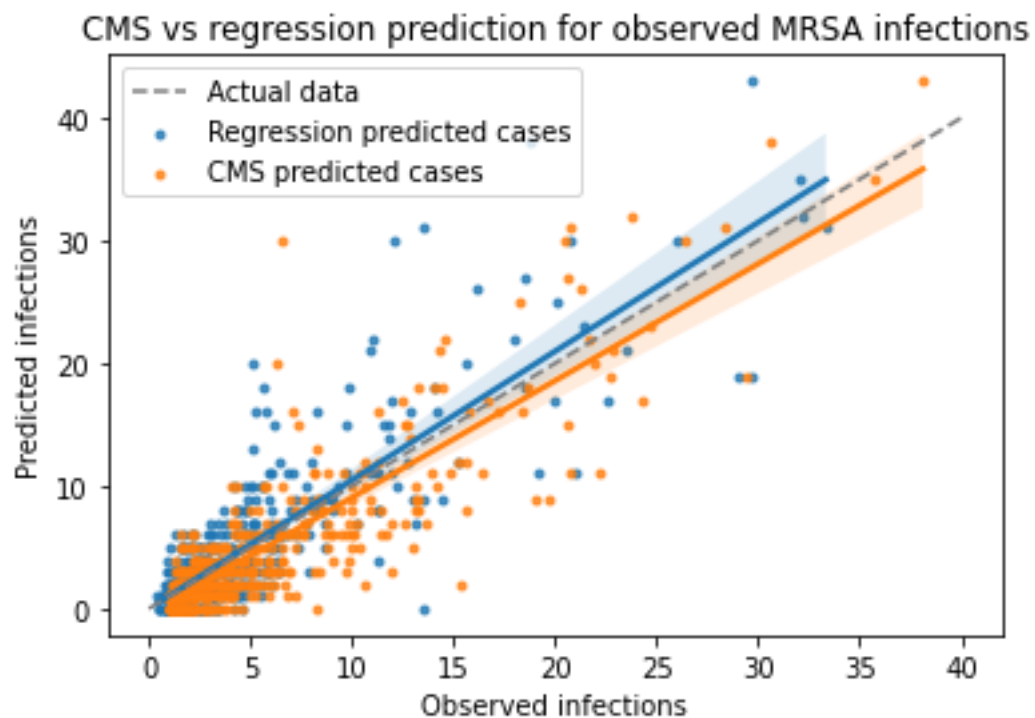
- A few key features dominate the model importance, all are HAI related

Number	Description
HAI_1	Central Line Associated Bloodstream Infection
HAI_2	Catheter Associated Urinary Tract Infections
HAI_3	SSI - Colon Surgery
HAI_4	SSI - Abdominal Hysterectomy
HAI_5	MRSA Bacteremia
HAI_6	Clostridium Difficile (C.Diff)



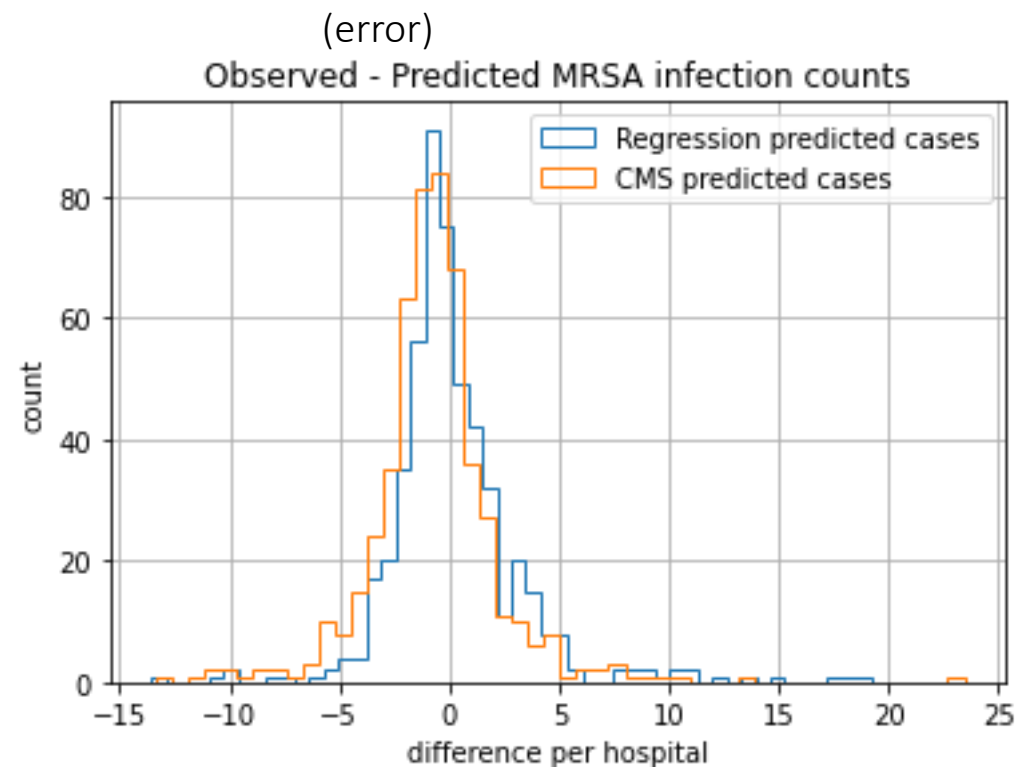
Outcomes

Comparison of my random forest regression prediction with CMS' prediction of observed MRSA cases for 2019 per facility



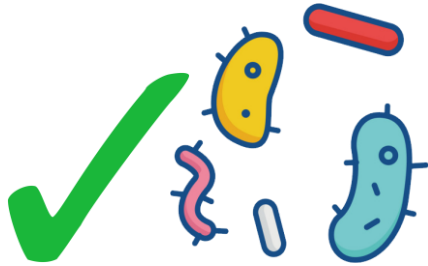
My regression MAE: 1.991

CMS MAE: 2.063



Conclusions

- It is possible to predict the number of MRSA infections expected with some error
 - Most important data for prediction is other HAI types
 - Patient survey results don't help predict MRSA infections
 - Sepsis care data don't help predict MRSA infections



HAI data



Patient Surveys



Sepsis treatment

Future Work

- Incorporate healthcare facility size (how many patients seen per year) and type of facility.
- Include historic data on infections
- Look more at geographic trends, especially with historic data
- Even more fun: hone in beyond the hospital aggregation level - what tests, procedures, and other data exists, ideally at a by-case detail level to help unravel this problem?

Thanks - Questions