STA 210 Final Project

Logan Godfrey and Krisztian Meszaros

Introduction and data

Hospitals are an important part of the health system of the United States. Finding ways to evaluate the impact that these health centers have on the health of their patients and their overall outcomes is important to understanding how we can better take care of patients with the resources that we have. There have been multiple ways that have been proposed to measure hospital quality, including the hospital standardized mortality ratio, which aims to categorize good and bad hospitals using the expected mortality and actual mortality of a hospital's patients given the mix of cases that they see (Wright and Shojania 2009). But adjustments to accurately control for case quantities are difficult to evaluate and may cause more harm than good. Although this method will likely be used because the data that it requires is easily accessible and cheap to gather, it may not be the best solution to measure hospital quality (Wright and Shojania 2009).

Some studies, using data related to hospital standardized mortality ratios, seek to understand the important factors that are associated with hospital quality. One such study analyzed the effect of hospital competition on overall quality (measured using hospital standardized mortality ratios) and found that competition is actually related negatively to overall quality as measured by this standard (Palangkaraya and Yong 2012). With these results in mind, it seems that more research and analysis could be helpful to the pursuit of safer and more effective hospitals. Our motivation primarily stems from a desire to research what factors are correlated with better hospitals in an effort to improve hospital quality and better patient care in the United States.

Our data is sourced from The Collection of Really Great, Interesting, Situated Datasets (CORGIS). A link to this project's website is provided here: https://corgis-edu.github.io/corgis/. The specific dataset that we used is the Hospitals CSV file listed under the CSV section of the website. A link to the specific page that contains the dataset is provided here: https://corgis-edu.github.io/corgis/csv/hospitals/. This data was initially collected from the Centers for Medicare & Medicaid Services, which is an official site of the U.S. government. Using this data, which is different from some of the other factors that have had their association with hospital quality assessed, we intend to research the following question: for hospitals in the mainland U.S., what factors are associated with the overall rating?

A data dictionary for the variables that we used from this dataset as is provided here:

- Facility.State - State where the hospital is located
- Facility.Type - Type of entity that operates the hospital. Is one of Government, Private, Proprietary, Church, or Unknown
- Rating.Overall - Single number rating of the hospital from 1 (lowest) to 5 (highest)
- Rating.XXXX - Rating of either mortality, safety, readmission, experience, effectiveness, timeliness, and imaging. Rating can take values of "above", "same", "below", or "unknown" when compared to the national standard of that category.
- Procedure.XXXX.Quality - Average quality to have a procedure for a heart attack, heart failure, pneumonia, or hip/knee condition. Rating can take values of "lower", "average", "better", or "unknown" when compared to the national standard of that category.
- Procedure.XXXX.Value - Average value to have a procedure for a heart attack, heart failure, pneumonia, or hip/knee condition. Rating can take values of "lower", "average", "higher", or "unknown" when compared to the national standard of that category.

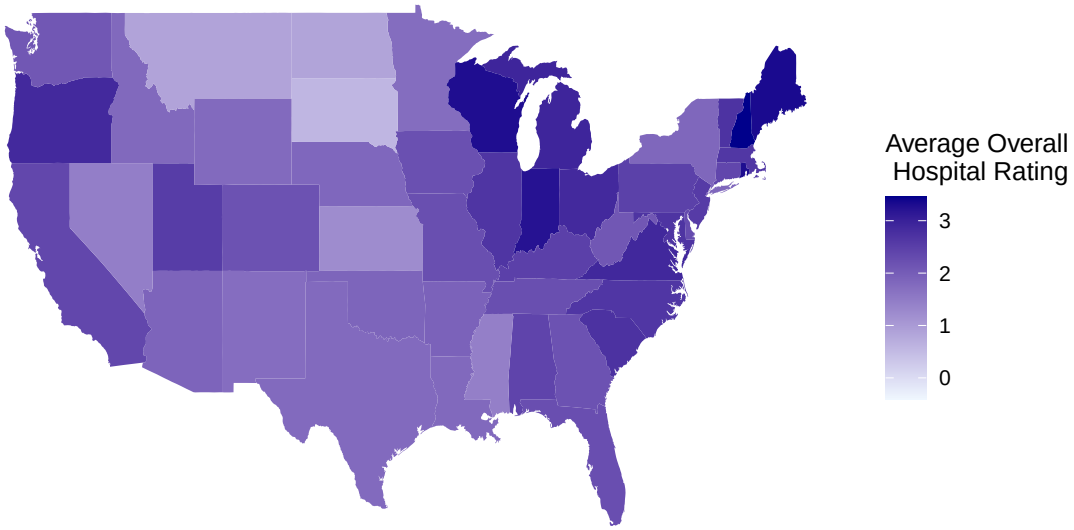
We also created variables for use in our analyses. The method that we used to create them is listed below:

- rating_avg - Average of all ratings given for mortality, safety, readmission, experience, effectiveness, timeliness, and imaging. The value above was given a 3, the value same was given a 2, the value below was given 1, and the value unknown was given NA. Values range from 1 to 3.
- quality_average - Average of all ratings given for procedure quality. The value better was given a 3, the value same was given a 2, the value below was given 1, and the value unknown was given NA. Values range from 1 to 3.
- value_average - Average of all ratings given for procedure value. The value higher was given a 3, the value average was given a 2, the value lower was given a 1, and the value unknown was given NA. Values range from 1 to 3.

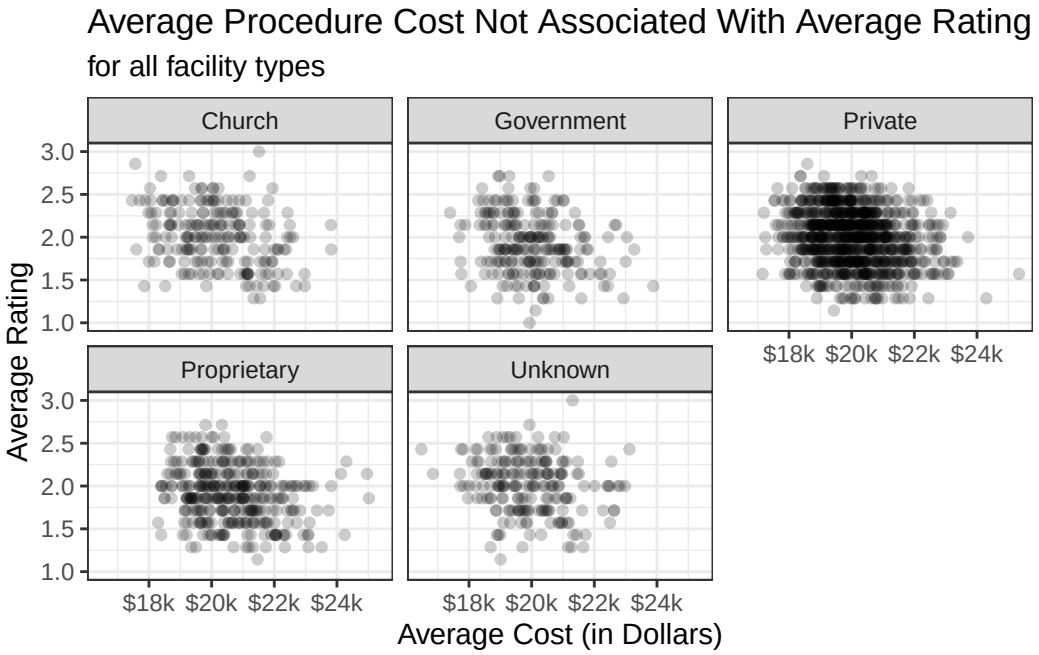
Our team also removed hospitals for which we did not have sufficient data to carry out our analyses.

We began with initial exploratory data analysis. To start, in an attempt to understand more about the distribution of hospitals around the country and how location is related to hospital quality, we graphed average overall hospital ratings by state. This graph is shown below:

Northeastern states tend to have higher average ratings



Next, we made a figure that examines the association between the average price of a hospitals procedures and the average rating across all different types of facilities.



Methodology

Our team chose to use an ordinal model for our research question. This is because our outcome that we are trying to predict, hospital rating, is an ordinal variable, with one being the worst rating and five being the best possible rating. Initially, we considered many of the variables in our dataset, including facility type and the average rating, average procedure quality, and average procedure value variables that we created. From there, we added the costs of procedures to the regression, and found that they had very little effect on hospital rating. We then broke average rating into its component ratings to see each of their effects on the overall rating, still adjusting for the other variables we originally used. We then considered an interaction term between average quality and average value, but found that this term had relatively little association with the outcome. Next, we assessed how our model fit the proportional odds assumption. We found that there is no reason to believe that the association of any of the variables that we used to model the overall rating for a hospital should vary depending on the level of rating that we are considering. Between any two levels, the effects of rating, quality, and value should be the same strength.

Results

Variable	Coefficient (Exponentiated)
Church Facility	Baseline
Government Facility	0.539
Private Facility	0.825
Proprietary Facility	0.580
Unknown Facility	0.924
Mortality Rating	4.958
Safety Rating	5.740
Readmission Rating	5.152
Experience Rating	4.629
Effectiveness Rating	1.468
Timeliness Rating	1.539
Imaging Rating	1.500
Quality Average Rating	29.777
Value Average Rating	0.708

Our model predicts that government, private, proprietary, and unknown hospital facilities would have less odds of being in the next highest overall rating level when compared to church hospital facilities while controlling for the other variables in our model. Additionally, our model predicts that government facilities are 0.539 times as likely, private facilities are 0.824 times as likely, proprietary facilities are 0.580 times as likely, and unknown facilities are 0.924

times as likely to be in the next highest overall rating level as church hospital facilities while controlling for the other variables.

Each individual facet that the hospital scores on were found to have varying associations with overall hospital score. While controlling for the other variables in the model, we predict that a hospital with one higher score in mortality would have 4.958 times the odds of being in the next overall score category compared to the lower-scored hospital. This prediction is similar for the categories of safety, readmission, and experience, with those having 5.740, 5.152, and 4.629 times the odds respectively of being in the next highest overall hospital rating group. The association that the variables effectiveness, timeliness, and imaging had on the outcome were much less pronounced. When controlling for the other variables in the model, we predict that having a one higher score in effectiveness, timeliness, and imaging will produce 1.468, 1.539, and 1.500 times the odds respectively of being in the next highest overall hospital rating group.

The average quality of procedures at the hospital facility is also predicted to have an association with the overall hospital score, as when controlling for all other variables in the model we predict that a hospital with a one higher procedure score will have 29.666 times the odds of being in the next highest overall hospital rating group. All of the variables assessed had associations with overall hospital rating, with a general trend being that having higher ratings in individual categories predicted higher odds of being in the next highest overall hospital rating category.

Discussion

Our analysis shows that there are some aspects of a hospital that are more predicted to have higher associaiton with a person's overall rating of a hospital than others. The average cost of procedures seems to have very little association with average rating. This would suggest that patients do not put much weight on the cost of the procedures when rating the hospital. The factors that were most important to patients were safety, readmission, and experience, while factors that were less important were effectiveness, timeliness, and imaging. It makes sense that a hospital that is unsafe or fails to properly readmit its patients will suffer in terms of rating, while some of the other factors are more negotiable.

While these findings do provide interesting insight, there are certainly limitations to this analysis. First of all, it is relatively unclear how the average hospital rating metric was collected, and whether it was administered in the same way for all of the hospitals. If the data for each hospital is an aggregate of responses from many different patients, it would have been useful to see the responses for each patient for each hospital. This would help in making our analysis more accurate and more insightful at the individual hospital level. Differences in the ways in which the ratings were collected could also introduce bias into the results. Additionally, the data contains a considerable amount of missingness in certain metrics, which is something that could be further looked into and better accounted for.

This preliminary analysis paves the way for further analysis into this area. One improvement on our analysis could be looking at patient-level data, and create better aggregate measures and categories to perform a more in-depth analysis. Additionally, further analysis could account for state-to-state differences (average individual income, for example) to better understand how hospitals compare to each other across states.

Works Cited

Palangkaraya, Alfons, and Jongsay Yong. 2012. "Effects of Competition on Hospital Quality: An Examination Using Hospital Administrative Data." *The European Journal of Health Economics* 14 (3): 415–29. https://doi.org/10.1007/s10198-012-0386-7.

Wright, J., and K. G Shojania. 2009. "Measuring the Quality of Hospital Care." *BMJ* 338 (mar18 2): b569–69. https://doi.org/10.1136/bmj.b569.