

# IDS 702 Final Project: Interpreting Hospital Ratings

Sydney Donati-Leach

12/12/2021

## Summary:

In order to understand and properly interpret why a hospital receives a certain star rating from the Centers for Medicaid and Medicare Services (CMS), one must take into account different quality measures. Mortality, safety, readmission, and patient experience are the quality measures that have the most impact on the rating. Additionally, the state in which the hospital is located has an impact on the star rating. The three states with the best hospitals are Ohio, Minnesota and Texas. The three states with the worst hospitals are Florida, New York, and U.S. territories.

## Introduction:

Every year, CMS rates every hospital registered with Medicare. The rating falls on a scale from one to five: one being the worst, five being the best. This “star rating” is derived from several different quality measures including mortality, safety, readmission, patient experience, timeliness, effectiveness of care, and efficient use of medical imaging. The measures selected come from a wide agreement among CMS, the hospital industry and public sector stakeholders.

These quality measures are evaluated in comparison to the national average and can be distinguished as above, below or the same as the national average. For example, a hospital could be ranked as below the national average for mortality, but above the national average for patient experience. Some hospitals may not report enough information on a quality measure for CMS to appropriately assign it a ranking. In this case, the quality measure will be listed as “Not Available”. CMS has stated the weight of the other quality measures will increase in this situation.

The goal of this report will be to interpret which of these quality measures have the strongest impact on the hospital star rating. Additionally, the hospital is put into one of ten different categories regarding its ownership. Therefore, the questions of interest for this report are the following:

- Which quality measures have the strongest impact on hospital rating?
- How does the hospital rating differ for the different types of hospital ownership (e.g., Government, Non-profit, For-profit, etc.)?
- Does the hospital rating differ by state? Which states differ the most from other states?

## Five Level Response:

To answer these questions, we will start by excluding state as the hierarchy and build a regular proportional odds model. After we have a model we are satisfied with, we will move to a hierarchical proportional odds model by including the state in which the hospital is located. Our response variable will be the hospital star rating, and the proportional odds model will have five levels (1 to 5). To look at the number of observations for each star rating, we can see the following distribution:

Table 1: Five Level Response

1	107
2	678
3	1761
4	939
5	82

There are a total of 3567 observations where each observation represents a hospital. This is a relatively normal distribution with the most amount of hospitals falling in the middle at 3, and the least amount of hospitals on the tail ends at 1 and 5.

## EDA:

Since each of the predictors in this data are factor variables, the only EDA we can do is through tables with the count of observations or the proportions. The first variable to look at is hospital ownership since this is a question of interest. As can be seen in the table below, there are 10 different categories of hospital ownership. The Tribal category only had 2 observations so we cannot take that into account when looking at trends. We can see that the trend does change as Physician owned hospitals have more hospitals with a 5 star rating than any of its other ratings. This is not the case for the other hospital ownership categories, so we will keep this in mind as we move onto model building.

Table 2: Hospital Ownership

	1	2	3	4	5
Government - Federal	0	5	7	4	0
Government - Hospital District or Authority	8	59	193	73	4
Government - Local	13	49	120	59	2
Government - State	8	11	18	7	1
Physician	0	3	6	4	7
Proprietary	26	177	296	111	7
Tribal	0	1	1	0	0
Voluntary non-profit - Church	4	56	144	91	14
Voluntary non-profit - Other	9	54	192	108	10
Voluntary non-profit - Private	39	263	784	482	37

There were two quality measures that were the most interesting and the tables of proportions can be seen below. First, if a hospital was ranked below the national average for mortality, 0% of those hospitals received a 5 star rating. Second, if a hospital was ranked below the national average for patient experience *or* it was not available, 0% of those hospitals received a 5 star rating. Readmission follows close behind where 0.12% of hospitals received a 5 star rating if this quality measure was rated below the national average. The remaining quality measures are safety, timeliness, effectiveness of care, and efficient imaging. If any of these quality measures were ranked below the national average, between 0.3% to 1.9% of hospitals were able to receive a 5 star rating. Based purely on this EDA, it appears the quality measures in this order have a decreasing significance on the star rating, but we will need to explore this further in model building.

Table 3: Mortality

	Above the national average	Below the national average	Not Available	Same as the national average
1	0.11	0.26	0.01	0.62
2	0.09	0.20	0.03	0.69
3	0.09	0.08	0.08	0.75
4	0.15	0.03	0.07	0.75
5	0.40	0.00	0.16	0.44

Table 4: Patient Experience

	Above the national average	Below the national average	Not Available	Same as the national average
1	0.00	0.95	0.01	0.04
2	0.05	0.72	0.03	0.20
3	0.20	0.24	0.13	0.43
4	0.66	0.07	0.04	0.24
5	0.88	0.00	0.00	0.12

One more point on EDA is that we will not be able to look at any interactions of variables due to the lack of observations when these quality measures are broken up.

## Model Building:

After taking into consideration everything we learned from EDA, we will start to build our proportional odds model. The first model will be simple and only contain the 2 quality measures we found the most interesting: mortality and patient experience. Moving to the second model, we will add readmission to the model. To test the significance of readmission, we can do an ANOVA Chi-squared test. Readmission is statistically significant, so we can move onto the next model with that included and test the next quality measure, safety, for significance. After repeating this process for all of the quality measures,

we will find that all are statistically significant, so we will want to include all of them in our model. After solidifying all the necessary quality measures in the model, the last step is adding hospital ownership, and this is statistically significant as well.

Based on what we learned from the ANOVA Chi-squared tests, our proportional odds model is as follows:

$$\log \left( \frac{Pr[y_i \leq j|x^i]}{Pr[y_i > j|x^i]} \right) = \beta_{0j} + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8}, \quad j = 1, \dots, J - 1$$

On the left side of the equation, we combine the first  $j$  star rating to form a single category and the remaining  $j$  star ratings to form a second category, where  $i$  represents hospitals. On the right side of the equation we have our predictors where  $x_{i1}$  is *owner*,  $x_{i2}$  is *mortality*,  $x_{i3}$  is *safety*,  $x_{i4}$  is *readmis*,  $x_{i5}$  is *patient\_exp*,  $x_{i6}$  is *effective*,  $x_{i7}$  is *timely*, and  $x_{i8}$  is *imaging*.

We will not do any interpretation of this model as this is not our final model. We still need to explore hierarchy based on state, and this will be covered in the next section.

### Three Level Response:

As stated before, the next step will be to include state as the hierarchy and build a hierarchical proportional odds model. However, we do not have enough observations to do this with a five level response. Therefore, we will reduce our response to only 3 levels. In this case, an original star rating of 1 or 2 would be mapped to a rating of 1, an original star rating of 3 would be mapped to 2, and an original star rating of 4 or 5 would be mapped to a 3. By collapsing the response to only 3 levels, there will be more observations present in each category for each state. An additional step we will take is collapse the U.S. territories of American Samoa, Guam, Puerto Rico, Virgin Islands, and Northern Mariana Islands into a state called "Others". These territories do not have many observations on their own, so combining them will allow us to evaluate their hospitals without being removed from the model.

After doing this, we can see the following distribution of observations:

Table 5: Three Level Response

1	785
2	1761
3	1021

### EDA:

When introducing a hierarchy, we must determine if there is varying intercept and/or varying slope. We can do this by looking at the count and proportions of observations for all the states. Based on the count of observations, it is clear that we will not be able to account for varying slope as there are less than 10 observations for many states. In addition, 7 states are missing observations when split across the 3 star ratings. However, when we look at the proportions of observations, the trend is not the same across states. For example, 89% of hospitals in Alaska have a 2 star rating. This differs greatly from D.C. in which only 14% of hospitals have a 2 star rating. Therefore, we will need to add varying intercept to our model.

### Modeling Building:

We will use our model from the five level response as our base model to build on. We can begin by adding state simply as a categorical variable, but this does not allow information to be shared across states. Therefore, we will fit a hierarchical model to explore the random effects that different states may contribute to the model. States are used as the only hierarchy in the model. Again, random slopes were not considered as the model would not have been able to converge which such few observations within many of the states.

Our final hierarchical proportional odds model is as follows:

$$\log \left( \frac{Pr[y_i \leq j|x^i]}{Pr[y_i > j|x^i]} \right) = \beta_{0j} + \gamma_i + X_i \beta, \quad j = 1, \dots, J - 1$$

All of the same variables from the previous proportional odds model are represented by the vector  $X$  in this simplified equation. The only addition is that we now have a random intercept of *state* represented by  $\gamma$ .

The model summary shown below will be the basis for all interpretations and answers to our questions of interest.

<i>Predictors</i>	<b>rate_new</b>						
	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>				
1 2	0.00	0.00 – 0.00	<0.001				
2 3	0.00	0.00 – 0.00	<0.001				
owner [Government - Hospital District or Authority]	1.51	0.43 – 5.28	0.518	readmis [Below the national average]	0.01	0.01 – 0.01	<0.001
				readmis [Not Available]	0.16	0.07 – 0.36	<0.001
owner [Government - Local]	1.16	0.33 – 4.10	0.820	readmis [Same as the national average]	0.10	0.08 – 0.14	<0.001
owner [Government - State]	0.81	0.19 – 3.46	0.778	patient_exp [Below the national average]	0.01	0.01 – 0.02	<0.001
owner [Physician]	1.56	0.29 – 8.50	0.608	patient_exp [Not Available]	0.08	0.05 – 0.12	<0.001
owner [Proprietary]	1.32	0.38 – 4.56	0.657	patient_exp [Same as the national average]	0.10	0.08 – 0.12	<0.001
owner [Tribal]	0.17	0.01 – 4.94	0.301	effective [Below the national average]	0.34	0.19 – 0.63	<0.001
owner [Voluntary non-profit - Church]	2.10	0.60 – 7.38	0.247	effective [Not Available]	0.87	0.44 – 1.72	0.690
owner [Voluntary non-profit - Other]	2.07	0.59 – 7.20	0.253	effective [Same as the national average]	0.93	0.57 – 1.51	0.761
owner [Voluntary non-profit - Private]	1.83	0.54 – 6.25	0.334	timely [Below the national average]	0.36	0.27 – 0.48	<0.001
mortality [Below the national average]	0.01	0.00 – 0.01	<0.001	timely [Not Available]	0.83	0.52 – 1.33	0.446
mortality [Not Available]	0.08	0.05 – 0.14	<0.001	timely [Same as the national average]	0.69	0.56 – 0.86	0.001
mortality [Same as the national average]	0.08	0.05 – 0.10	<0.001	imaging [Below the national average]	0.33	0.22 – 0.49	<0.001
safety [Below the national average]	0.01	0.01 – 0.02	<0.001	imaging [Not Available]	0.66	0.45 – 0.98	0.039
safety [Not Available]	0.18	0.12 – 0.25	<0.001	imaging [Same as the national average]	0.86	0.62 – 1.18	0.335
safety [Same as the national average]	0.12	0.09 – 0.15	<0.001				
<b>Random Effects</b>							
$\sigma^2$					3.29		
$\tau_{00}$ State					0.05		
ICC					0.02		
N State					51		

First, it is important to note that for our response of 1|2 and 2|3 the coefficients are not actually zero, but are extremely small. Secondly, our baseline in the model is a Government Federally owned hospital, with all of the quality measures ranked above the national average. The Voluntary non-profit hospitals all have coefficients larger than 1, signifying that the odds of these hospitals receiving a rating towards a 5 rather than towards a 1 is roughly two times greater than a Federally owned hospital. However, this is not statistically significant due to the p-values being greater than 0.05.

Moving onto our quality measures, all of these coefficients are less than one which is reasonable given that our baseline hospital had all its quality measures above the national average. Additionally, almost all of the variables are significant. For 4 variables (mortality, safety, readmission, and patient experience) we see the exact same coefficient of 0.01 when the quality

measure is ranked below the national average. This means that if a hospital receives a ranking below the national average in one of these quality measures, it is 99% less likely to get a star rating towards a 5.

For the other 3 quality measures, (timeliness, effectiveness of care, and efficient imaging), the coefficients are larger when it is ranked below the national average. For example, if a hospital receives a ranking below the national average for timely care, it is only 64% less likely to get a star rating towards a 5. This shows that these 3 quality measures have less of an impact on the overall star rating, as compared to the other 4 quality measures.

If we look at the random effects of this model, the standard deviation of star ratings between states is 0.23. We can also evaluate which states have hospitals with star ratings necessarily worse or better than the other states' hospitals. The three states with the best hospitals are Ohio, Minnesota and Texas. The three states with the worst hospitals are Florida, New York, and the U.S. territories we combined into "Others." Figure 1 in the appendix shows a full map of the United States designating which states are in the top half in the country verses the bottom half.

Finally, regarding model assessment, we are not able to create binned plots of residuals since all of our variables are factor variables. Additionally, we cannot look at a binned residual plot of the overall expected values. The cumulative link mixed model (clmm) that was used for this hierarchical proportional odds model creates cumulative predicted probabilities from the fitted values rather than exact predicted probabilities.

## Conclusions:

We were able to answer all of our questions of interest using the hierarchical proportional odds model. The answers are as follows:

- Which quality measures have the strongest impact on hospital rating?
  - Mortality, safety, readmission, and patient experience.
- How does the hospital rating differ for the different types of hospital ownership (e.g., Government, Non-profit, For-profit, etc.)?
  - The ratings do differ by hospital ownership, and voluntary non-profit hospitals may potentially outperform the other hospitals, but the lack of statistical significance makes this an inconclusive statement.
- Does the hospital rating differ by state? Which states differ the most from other states?
  - Yes. The three states with the best hospitals are Ohio, Minnesota and Texas. The three states with the worst hospitals are Florida, New York, and the U.S. territories we combined into "Others".

The limitations to this report came down to the data available, and the limited number of observations in certain categories. If there were more observations, we would have been able to explore interactions, and we would not have had to switch from a 5 level response to a 3 level response. Along those same lines, we might have been able to explore random slopes and determine if there is any variation within states. The future potential of this project would be to add in more years of data to gain more observations.

Figure 1:

