Uncovering value in hospital patient surveys

Alex Austin

July 25, 2021

New York City Data Science Academy

Challenges to the healthcare business

Providers can't charge higher rates. Payers can't increase premiums. [...] To uncover value and grow business, stakeholders must start to chip away at the cost of care.

- Mitchell Morris, MD, Optum

Main stakeholders:

- insurance companies
- hospitals

Research proposal

Project outline: We link patient survey data to hospital performance metrics to identify means of reducing costs (by improving the quality of healthcare).

Data source: The Centers for Medicare & Medicaid Services (CMS) Medicare provider data (*https://data.cms.gov/provider-data/*). In particular, the datasets *Patient survey (HCAHPS) - Hospital*, *Unplanned Hospital Visits -Hospital*, and *Hospital-Acquired Condition Reduction Program*.

Response variables

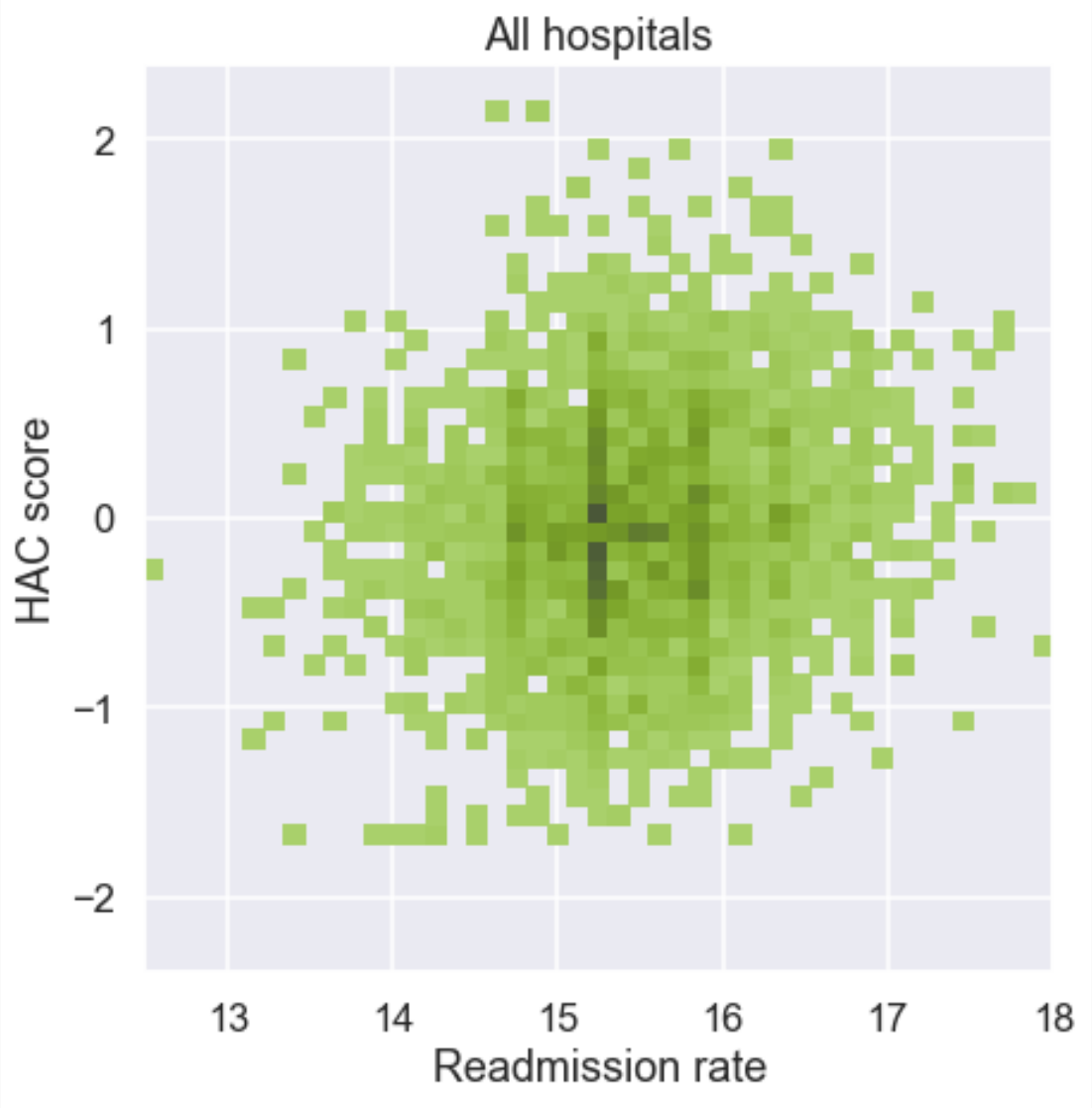
(1) Rate of readmission within 30 days of being discharged from a hospital stay.

(2) Total Hospital-Acquired Condition (HAC) Score (signed standard deviations from mean).

Hazards to stakeholders:

- Increased costs/claims payments.
- Damage to reputation.
- Reduced funding.
- Greater likelihood of malpractice lawsuit.

Little relationship between targets



Predictors

Amalgamated at the hospital level from patient surveys, they are the linear scores attributed to

- Cleanliness
- Nurse communication
- Doctor communication
- Staff responsiveness
- Communication about medicine
- Discharge information
- Care transition
- Overall hospital rating
- Quietness
- Recommend hospital

Readmission rate: best subsets regression

OLS Regression Results

Dep. Variable:	y	R-squared:	0.091
Model:	OLS	Adj. R-squared:	0.088
Method:	Least Squares	F-statistic:	33.81
Date:	Sun, 25 Jul 2021	Prob (F-statistic):	2.58e-63
Time:	13:37:10	Log-Likelihood:	-3610.9
No. Observations:	3384	AIC:	7244.
Df Residuals:	3373	BIC:	7311.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	20.6760	0.617	33.484	0.000	19.465	21.887
x1	-0.0150	0.005	-3.141	0.002	-0.024	-0.006
x2	-0.0090	0.012	-0.734	0.463	-0.033	0.015
x3	-0.0164	0.009	-1.743	0.081	-0.035	0.002
x4	-0.0126	0.006	-2.181	0.029	-0.024	-0.001
x5	0.0206	0.005	3.818	0.000	0.010	0.031
x6	-0.0469	0.005	-8.998	0.000	-0.057	-0.037
x7	0.0407	0.011	3.763	0.000	0.019	0.062
x8	0.0392	0.015	2.539	0.011	0.009	0.069
x9	-0.0079	0.003	-2.309	0.021	-0.015	-0.001
x10	-0.0470	0.009	-5.009	0.000	-0.065	-0.029

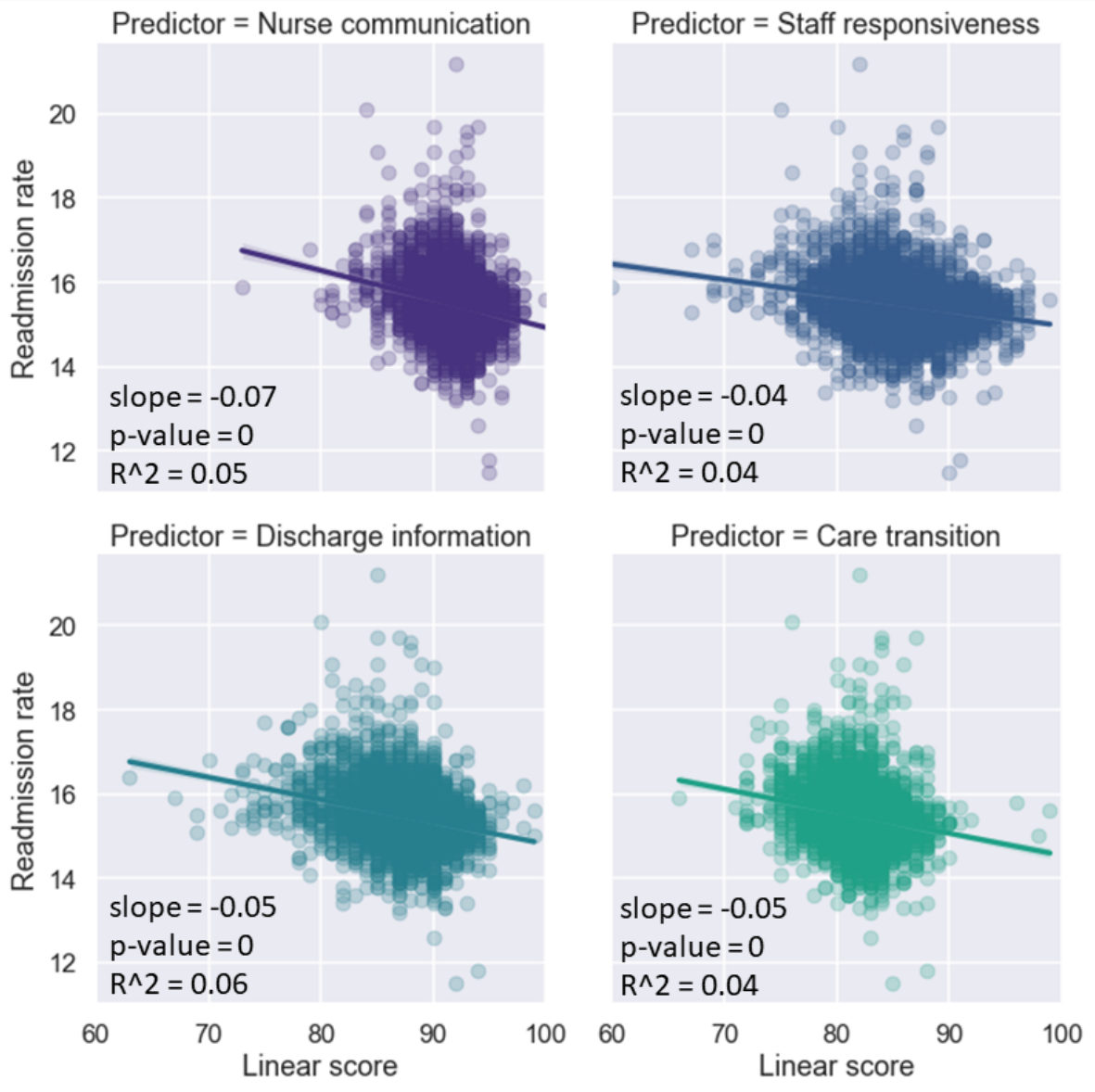
Omnibus:	537.283	Durbin-Watson:	1.803
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2526.924
Skew:	0.688	Prob(JB):	0.00
Kurtosis:	7.004	Cond. No.	1.39e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.39e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

Readmission rate: correlations

Discharge information	-0.262240
Recommend hospital	-0.232840
Nurse communication	-0.226153
Overall hospital rating	-0.225432
Staff responsiveness	-0.212882
Care transition	-0.203613
Doctor communication	-0.197653
Cleanliness	-0.193883
Communication about medicine	-0.167898
Quietness	-0.157066

Readmission rate: concentrate on these predictors



Total HAC score: correlations

Staff responsiveness	-0.227254
Cleanliness	-0.211575
Quietness	-0.202352
Communication about medicine	-0.191847
Care transition	-0.180225
Nurse communication	-0.179704
Overall hospital rating	-0.173021
Doctor communication	-0.165452
Discharge information	-0.143240
Recommend hospital	-0.138713

Comparing strongest correlations for the two targets

Readmission rate	**Total HAC score**
Discharge information	Staff responsiveness
Recommend hospital	Cleanliness
Nurse communication	Quietness
Overall hospital rating	Communication about medicine

The intersection is empty.

Conclusions and recommendations

- The regressions were statistically significant.
- The effect sizes were small.
- Results were qualitatively interesting.

Relatively high readmission rate: try to improve discharge information and nurse communication.

In the data examined, each one point increase in the discharge information score corresponds to a 0.05% decrease in the 30 day readmission rate.

Relatively high total HAC score: try to improve staff responsiveness and cleanliness.

In the data examined, each one point increase in the staff responsiveness score corresponds to a 0.03 s.d. decrease in the total HAC score.

Future work

Improve

- Resolve issue of mismatched collection period dates.
- Better understand the patient survey process and timeline.
- Adapt regression methods to account for multicollinearity.
- Refine approach to address non-constant variance.

Extend

- Segment the data prior to analysis.
- Consider other response variables, such as from the *Complications and Deaths* dataset.

Replace

- Patient level data that would allow for a machine learning approach to predicting which patients are at high risk of readmission.

Thank you.

Any questions?