Data Immersion
Exercise 6.1
Jennifer Ignacio

Data Source:

Hospital Readmission Reductions Program

- https://data.cms.gov/provider-data/dataset/9n3s-kdb3#data-dictionary

Hospital General Information

- https://data.cms.gov/provider-data/dataset/xubh-q36u#data-table

Hospital Value-Based Purchasing (HVBP) - Clinical Outcomes

- https://data.cms.gov/provider-data/dataset/pudb-wetr

Summary

The Hospital Readmissions Reduction Program, Hospital General Information, and Hospital Value-Based Purchasing (HVBP) - Clinical Outcomes Domain Scores are datasets are owned by CMS (Centers for Medicare and Medicaid Services), compiled from contributions by both public and private stakeholders including organizations like The Joint Commission (TJC), Partnership for Quality Measurement, and the Agency for Healthcare Research and Quality (AHRQ). CMS ensures the trustworthiness of this data through a robust validation process that rigorously checks against multiple sources such as claims data, surveys, and administrative records. This meticulous approach guarantees the accuracy and reliability necessary for informed decision-making in healthcare policy and management. These datasets are also timely as they are updated every fiscal year. I will be using the most recent 2024 datasets in my analysis.

Data Choice:

Hospital Readmission Reductions Program

I plan to use this dataset as the primary data source and integrate additional datasets to analyze how hospital characteristics impact readmission rates. By examining factors such as hospital ownership (private vs. public), emergency service support, and performance benchmarks, I aim to determine their influence on readmissions. While patient health is a significant factor, it is crucial to identify hospital-level contributions to readmission likelihood to enhance quality of care and achieve financial and resource efficiency. This integrated analysis will provide actionable insights to help hospitals reduce readmissions, optimize resource allocation, and improve overall patient outcomes.

Hospital General Information

I plan to use this dataset to supplement the readmissions dataset, as it contains important categorical information such as hospital type, ownership, and ratings. I will join these datasets using Facility ID and Facility Name as keys to provide a comprehensive analysis of how particular hospital characteristics affect readmission rates.

Hospital Value-Based Purchasing (HVBP) - Clinical Outcomes

I plan to use this dataset to supplement the readmissions dataset, as it contains crucial numerical information such as achievement thresholds, historical baselines, and performance benchmarks. It also includes scores, ratings, and improvement points. I will focus primarily on performance benchmarks, baseline rates, and performance rates to analyze whether continuous outperformance relative to historical baselines or peer benchmarks impacts readmission rates. I will merge these datasets using Facility ID and Facility Name as keys.

Data Profile:

Data Profile for datasets
Jupyter Notebook for profiling the data

Data Cleaning:

Jupyter Notebook for profiling the data

Data Consistency:

Consistency Check for datasets

Limitations and Ethical Considerations:

Timeliness: All of the datasets were updated this year. However, the Hospital Readmission Reductions Program datset's data was collected from 2019 to 2022 which makes this data about , 2 years old. Although still recent, wish a gap of two years between when the data was collected and when it became accessible to the public things could have changed, in addition to the ways the variables will interact based.

Data Accuracy: Due to the number of missing values, it was necessary to either drop rows/columns or impute values to clean the dataset. This reduces the overall sample size and can interfere with data accuracy, as imputation can make the data less reliable and potentially affect the analysis. (No more than 43% of the original sample size was removed from the data, with data imputation at 35% of the original data)

Suppression: Many data points were null or missing due to suppression to protect the identity of patients involved in data collection. This suppression occurs when there are too few patients, jeopardizing the principles of Personally Identifiable Information (PII) in data handling. Consequently, values had to be imputed or rows/columns with missing data had to be deleted, impacting the dataset's completeness.

Impact on Hospital Reputation: Analyzing and publishing data on hospital performance can reveal deficiencies in certain hospitals, potentially damaging their reputation. It is essential to present findings responsibly, considering the broader context and the limitations of the data.

Questions:

Impact of Hospital Characteristics:

- How do different hospital characteristics (e.g., type, ownership) impact readmission rates?

Readmissions and Performance Metrics:

- How do hospitals that meet or exceed excess readmissions thresholds differ in their overall ratings and other performance metrics?

Targeting Improvement Factors:

- For hospitals with high readmission rates, what common factors (e.g., count of safety measures, hospital type) can be targeted for improvement?
- Is there a correlation between hospitals that promote EHR interoperability and higher hospital overall ratings?

Condition-Specific Strategies:

- How do hospitals' performance comparisons across different conditions (e.g., AMI, HF, PN, COPD) inform condition-specific improvement strategies?

Regional Performance Analysis:

- Are there specific regions where hospitals consistently underperform or overperform in terms of readmission rates and overall ratings?