Financial Performance and Charity Care Analysis in Healthcare Facilities: Evaluation of Costs, Revenues, and Provider Characteristics

George Mason University
AIT-580-004 | Prof. Dr. Kazi Lutful Kabir

Praneet Reddy
George Mason
University
Fairfax, Virgina
schinth6@.gmu.edu

Abstract— This article examines the financial performance of healthcare institutions and charity care through the lens of primary and secondary variables such as cost, revenue, and operational indicators. Python, R, and MySQL tools are used to clean and find trends in healthcare providers' total expenditures, net income, patient revenues, and charity care. Data visualisation and statistical modelling are used to uncover key parameters such as bed count, total discharges, and Medicare utilization which in turn influence financial performance and service quality. The findings give meaningful insights into the operational and financial dynamics of the health care industry.

Key words: Healthcare Finance, Charity Care, Data Analysis, Revenue Optimization, Healthcare Operations, MySQL, R, Python

I. Introduction

The financial performance of healthcare institutions is one of the most critical aspects in modern healthcare systems, since it shows their ability to deliver good quality care while remaining economically sustainable. Hospitals face challenges in increasing costs related to operations, fluctuating volumes of patients, and the delivery of charity care. Therefore, it becomes very important to understand which factors will affect the income and expenses of hospitals for improvement in their financial strategies for better care quality.[1]

This following questions provide an insight with their answers related to financial and operational dynamics in hospitals. These questions are vital for understanding the data and extracting valuable information from it.

How do different types of hospitals affect their income and expenses?
Comparing financial metrics across hospital types by CCN facility type and service level helps reveal patterns in cost and revenue structures, thereby providing a general idea of the budget to be made.

Is there a link between the number of hospital beds and patient volume, and how does this affect total revenue?
Correlation analysis is used to examine relationships between bed count, patient volume, and total revenue, highlighting how effectively patients are handled in

terms of numbers.

How do hospitals manage charity care while maintaining financial stability?
Analysis of charity care costs as a percentage of net income and patient revenue is done. This identifies the right approaches to balance financial obligations and charity care for community support.

How does the severity of care provided (CCN Facility Type) impact hospital operating costs and patient outcomes?
Comparative analysis of operating costs and patient discharge data is done. This helps in assessing how care complexity influences the expenses and outcomes of the healthcare unit helping in managing the budget provided and estimation for the future.

Data cleansing, visualization, and statistical modeling are carried out using data processing tools from Python, R, and MySQL. The methods mentioned are then applied in analyzing hospital performance metrics. Results give practical insights to hospitals on how to enhance financial stability without compromising high-quality care standards. This contributes to the knowledge base in healthcare economics and management using a data-centric methodology, which provides important information to decision-makers in the healthcare sector.
Related Work:

Different methodologies have been applied regarding the analysis of financial performance and healthcare outcomes in the framework of hospital costs and care. The scoping review of Dubas-Jakóbczyk et al. (2020) explored the financial performance of hospitals and its association with quality of care. According to them, better quality care is generally provided by those hospitals which have better financial stability, and thus good financial health is important for optimal patient outcomes.

II. Related Work

Relatedly, Gafford (2020)[2] examined the impact of Medicaid expansion on hospital financial performance and healthcare outcomes using 69 related papers. Using a difference-in-differences regression model, Gafford found positive impacts from Medicaid expansion on the efficiency of hospitals, giving a clear framework for research on external policy changes and performance.

Marques and Alves (2023) performed a review of methodologies concerning hospital costing and highlighted the different approaches followed by hospitals in managing and allocating costs. This research underlines the importance of accurate costing techniques, which are essential in understanding financial dynamics within hospitals.

Further, Cohen and O'Brien (2023) tested how the Affordable Care Act (ACA) influenced hospital financial management and efficiency. Indeed, their research showed significant changes in cost structures and revenue generation following the implementation of the ACA. This study provides a critical context in which to analyze changes in hospital financial performance over time, particularly in response to large-scale policy interventions. Such insights are applicable in the present analysis, which aims to identify trends in hospital costs and care quality within a changing healthcare environment.

III. Methodology

III.1. Dataset Description

The dataset used in this study includes records related to hospital costs and care, totaling 4,233 records with 14 attributes. The dataset contains various attributes like provider CCN, CCN facility type, CCN category, number of beds, total costs, cost of charity care, total income, net income, net patient revenue, total days and

discharges for Title XVIII, and fiscal year details. The target variable is the hospital cost, continuous, representing the total cost each hospital incurred in rendering health care services.

Data Type	Columns
Nominal	Provider CCN
Ordinal	CCN Facility Type, CCN category
Ratio	Number of Beds, Total Costs, Cost of Charity Care, Total Income, Net Income, Net Patient Revenue, Total Days Title XVIII, Total Discharges Title XVIII
Interval	Fiscal Year Begin Date, Fiscal Year End Date

III.2. Data Preprocessing

The preprocessing of the dataset was done using Python and R codes. This process focused on handling missing values, transforming categorical data, and scaling numerical data. The presence of Null values in the data required a cleaning process that deleted those rows, so that only complete records remained for analysis. In the code, no explicit transformation to numerical values was carried out for categorical features such as ‘CCN Facility Type', and 'CCN category'. Continuous numeric data, such as 'Total Costs' and 'Net Income', were prepared for analysis by scaling so that these variables, having different ranges and units, would not disproportionately affect the model. After cleaning and transformation, the data became ready for further analysis and modeling.

III.3. Tools and Platforms

MySQL was used to store and query the dataset on hospitals. SQL queries were executed to fetch important data, like determining the top 5 providers based on net income, calculating the average cost of charity care, and analyzing the impact of hospital type on revenue. R and Python were used for data analysis and visualization. R was used to generate histograms and bar plots to analyze the distribution of features, while Python was used primarily for data cleaning. Filtering of the dataset was done, missing values were removed, and then the cleaned data was saved for further analysis.[3]

IV. Data Analysis and Results

EDA was done using R and Python to explore the dataset. Visualizations such as bar plots and histograms were used to understand the relationship between features like total costs, income, and hospital types. The analysis found that there are correlations among the key variables, including how the number of beds and facility type affect financial performance. Insights were drawn on how charity care and revenue are affecting the net income of a hospital. This finding pinpoints the need for managing various financial factors to maintain stability.

V. Linear Regression Model

A linear regression was conducted on the relationship between total costs and total income in health facilities. The correlation was significant, showing that as the income goes up, so do the total costs. Quantitatively, this is determined by the slope coefficient in the regression equation, which stands for the rate of changes in the costs relative to the fluctuation of income. The scatter plot below confirmed this positive association, as it indicated a clear upward trend. These findings have important implications for financial planning and resource allocation in the healthcare sector, with important pointers toward strategic budgeting based on income levels.

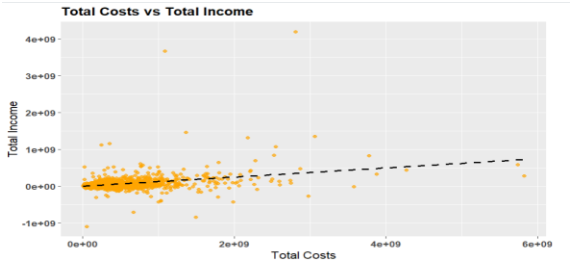
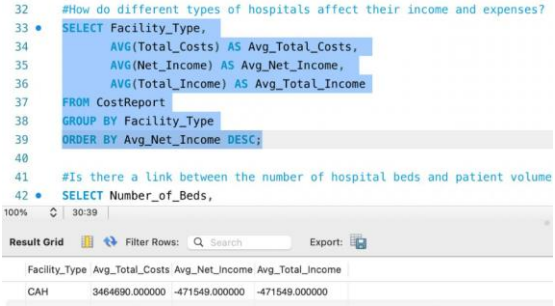


Fig: Total Cost vs Total Income

VI. Data Exploration using SQL

Exploration of SQL entails selecting unique combination of attributes, querying and analyzing data to uncover significant patterns. By applying few SQL commands, such as `SELECT`, `JOIN` or several functions of aggregate types `SUM()`, `AVG()`, users are able to select and sort the records in several tables. Data manipulation also supports by standard SQL, Transformations and sub query in database cleans and format the data for particular deep analysis. It's a must-have in the process of searching for trends or any data-based decisions.



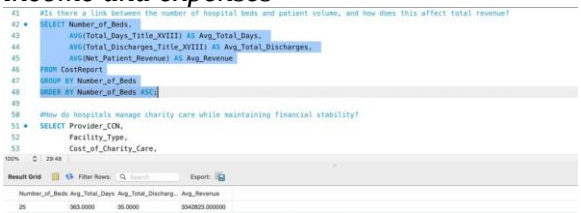
```
#How do different types of hospitals affect their income and expenses?
SELECT Facility_Type,
       AVG(Total_Costs) AS Avg_Total_Costs,
       AVG(Net_Income) AS Avg_Net_Income,
       AVG(Total_Income) AS Avg_Total_Income
FROM CostReport
GROUP BY Facility_Type
ORDER BY Avg_Net_Income DESC;

#Is there a link between the number of hospital beds and patient volume
SELECT Number_of_Beds,
```

Result Grid | Filter Rows: | Export:

Facility_Type	Avg_Total_Costs	Avg_Net_Income	Avg_Total_Income
CAH	3464690.000000	-471549.000000	-471549.000000

Fig: Different types of hospitals affect their income and expenses



```
#Is there a link between the number of hospital beds and patient volume, and how does this affect total revenue?
SELECT Number_of_Beds,
       AVG(Total_Days_Title_XVIII) AS Avg_Total_Days,
       AVG(Total_Discharges_Title_XVIII) AS Avg_Total_Discharges,
       AVG(Net_Patient_Revenue) AS Avg_Revenue
FROM CostReport
GROUP BY Number_of_Beds
ORDER BY Number_of_Beds ASC;

#How do hospitals manage charity care while maintaining financial stability?
SELECT Provider_CCN,
       Facility_Type,
       Cost_of_Charity_Care,
```

Fig: Describes the link between hospital bed and patient volume

VII. Discussion

How do different types of hospitals affect their income and expenses?

This question aims at knowing whether community, specialized or general hospitals post high or low income and or expenditure. Therefore, using the hospital data set, to compare Total.Income and Total.Costs by the categories that are based on the CCN.Facility.Type, we can apply boxplots. With the help of the visualization, it is easier to determine whether some kinds of hospitals tend to bring more income or lead to more costs. High variability of the income or costs can be identified with the help of the wide range in the boxplot while the low variability is associated with the narrow range of that hospital type.

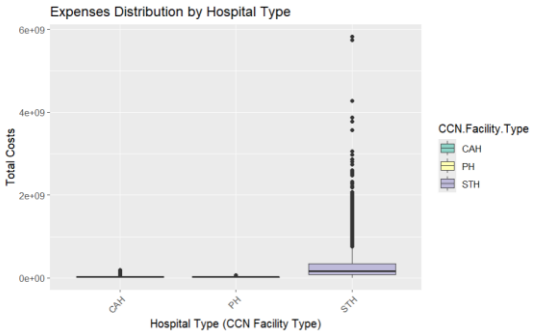


Fig: Expenses Distribution by Hospitals

Is there a link between the number of hospital beds and patient volume, and how does this affect total revenue?

This question seeks to assess the following; Do larger hospitals, with more beds admit and or attend to, more patients and consequently earn higher total revenues? Discharging a total of (Total.Discharges.Title.XVIII) with (Number.of.Beds) number of beds and generating an operating income of (Net.Patient.Revenue), scatterplots signify any relationship between the amount of capacity, volumes, and revenues. In these plots, if the values are inclined up, it will mean that the hospitals with larger bed capacities emit more volumes, thereby projecting more revenues; in contrast, low inclined values or straight lines point that the bed capacity should not determine volumes or revenues.

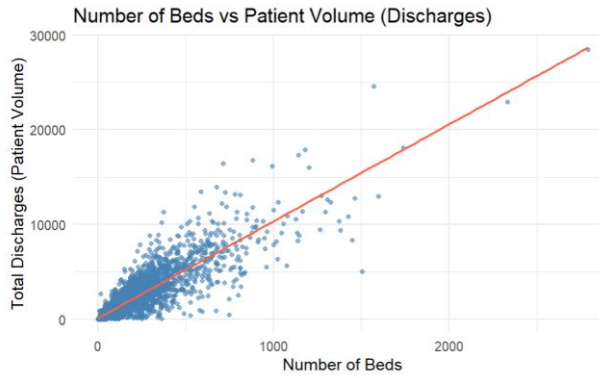


Fig: Number of beds vs patient Volume(Discharges)

How do hospitals manage charity care while maintaining financial stability?

This question also concerns the proportions of charity care and financial sustainability. Thus, costs of charity care, as shown in Cost.of.Charity.Care, are compared to hospitals' Net.Income to uncover the existence of a relationship between the level of charity care expenditure and financial results. Through plotting of these variables in a scatter plot, one is able to identify if Hospitals with high levels of charity care have poor financial stability. If the plot demonstrates the negative trend then it would mean that higher the charity care cost lower will be the net income of the hospital and therefore the insight should be that hospitals should act wisely while operating on charitable services.[2][4]

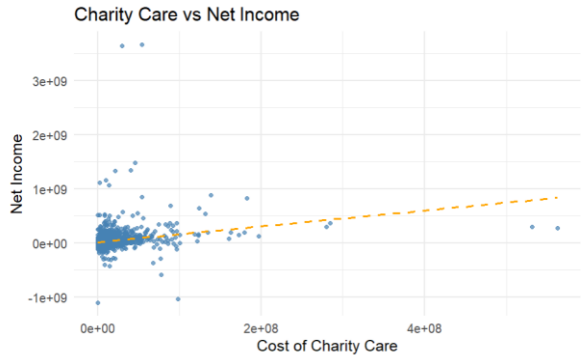


Fig: Charity Care vs Net Income

How does the severity of care given (CCN Facility Type) impact hospital operating costs and patient outcomes?

This question arises out of hypothesising that hospitals with a higher level of CCN.Facility.Type, which suggests more complicated or serious care, have both greater operating expenses and different patient outcomes. To analyze how Total.Costs and Total.Discharges.Title.XVIII are related to other facility types and if hospitals with superior-defined levels of care bear more significant costs and if patient results (discharge rates) are a consequence of these costs, boxplots are utilized to compare Total.Costs and Total.Discharges.Title.XVIII to other facility types. If analysis of the boxplots indicates that there are more severe care types with higher costs and the patients' discharge rates, it indicates that there is a likelihood of increased costs in the management of these treatments on the health of the patients.

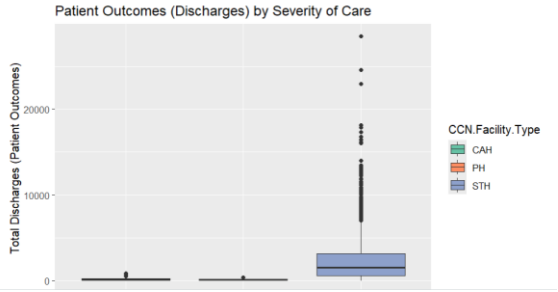


Fig: Patient Outcomes (Discharged) by Severity of Care

Univariate Analysis:

Nominal Data:

Nominal data refers to data that fall under categories that are not arranged in any order or ranked. In our data set, Provider.CCN is nominal by definition. The variable that is represented through this particular boxplot is the frequency counts of a provider code, which means that in the dataset when a particular code has repeated itself how often has it been repeated. This is nominal data, therefore, this involves presentation of the categories' frequencies, examination of extreme values, and spread of the categories. From the boxplot, it is clear that any outlier, that is any value that significantly falls outside the normal range of density of values, is presented by an asterisk.

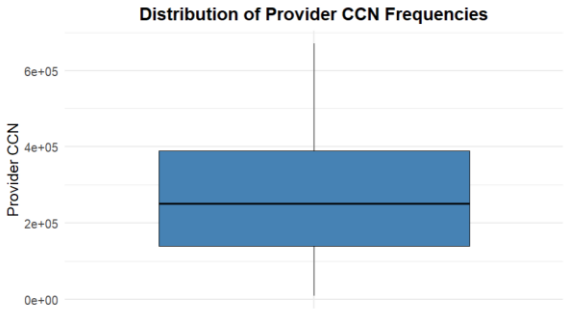


Fig: Distribution of Provider CNN Frequencies

Ordinal Data:

Ordinal data so has an order; however, the difference between the categories is not necessarily equal. The CCN.Facility.Type variable is ordinal data in the dataset because the variable has ranked the hospitals according to their type. A bar plot of CCN.Facility.Type can then be used to present the frequency distribution of hospital facilities. Every bar refers to the different facility types where it plots the number of hospitals for each type. This makes it possible to quickly familiarize with the basic facility types, and their share in the sample data set. While the categories are quite ologically ordered, plot is used to indicate how often and where these ordered categories occur.

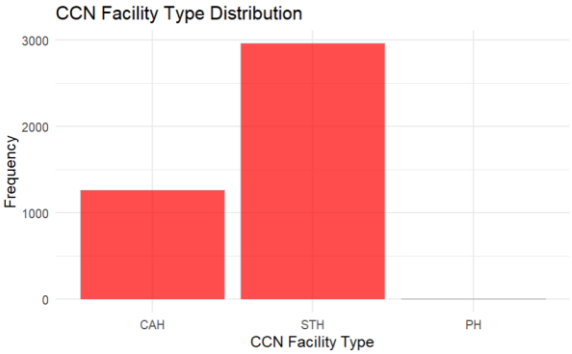


Fig: CNN Facility Type Distribution

Interval Data:

On our interval data, the numbers are numeric because the distances of one value from another are important, although there is no origin point. Total.Days.Title.XVIII, concerning the total days of Title XVIII (Medicare), is an interval variable in our dataset. The histogram below depicts this data in terms of the number of service day ranges that are present. Hence, plotting the histogram shown below will enable us determine if the data follow a normal distribution, identify outliers and measure skewness, and finally measure the variability by observing the range of days through which hospitals have provided under Medicare. The distribution of days tells information concerning how stretched out or compacted common hospital stay extents for the Medicare patient are.

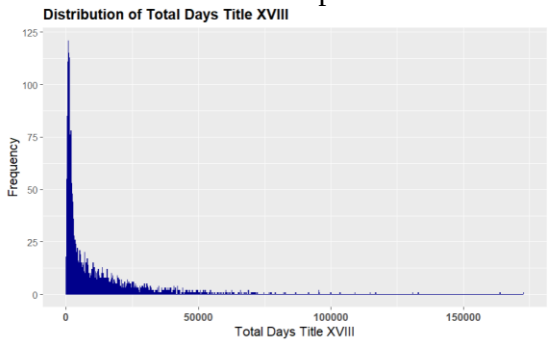
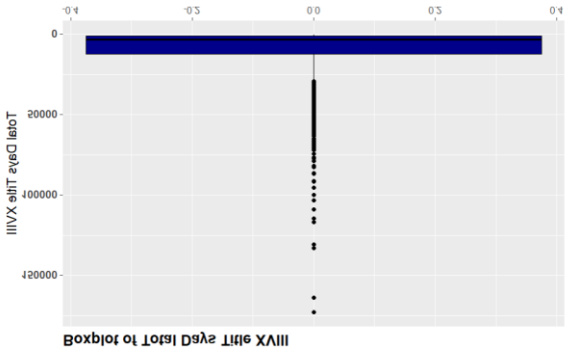


Fig: Distribution of Total Days Title XVIII

Ratio

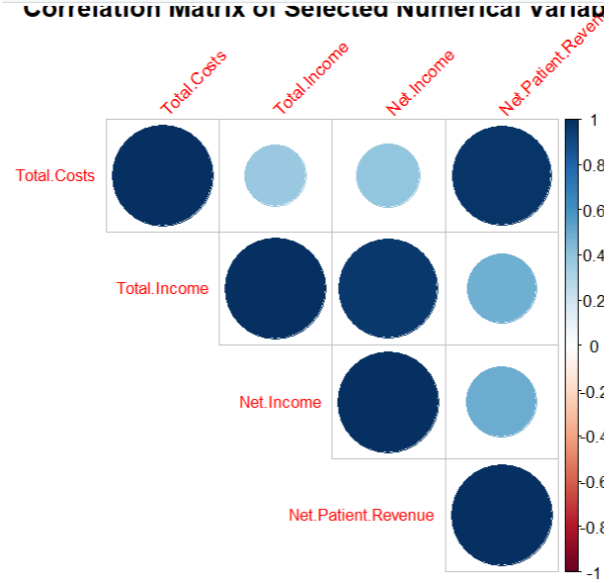
Similar to interval data, ratio data is not affected by the origin and the zero point is true and the value between two ratios is significant. Despite the fact that Total.Days.Title.XVIII was analyzed as interval level data earlier, it is possible to view it as ratio level variable in case of measuring the days spent in hospitals. The boxplot of this variable presents the dispersion, median and quartiles and possible outliers. An important idea of the use of ratio data is that one can obtain meaningful ratios, such as one hospital served twice as many service days as another. Boxplot graph makes it easier to compare the variability of variable total service days and to decide which hospital's data is excessively different to the rest. This also helps in finding out the distribution of the whole picture and to know where most of the hospitals are in the context of service days and possibly costs or anything at all.



Multivariate Analysis:

As shown on the correlation matrix below, most of the numerical variables have a moderate positive correlation.

In investigating the correlation between many numerical variables, the first activity is to construct a correlation matrix. Here we are looking at the relationship between total reported costs, total reported revenues, net income, and net patient revenue. To do this we need to use the cor() function which shows how close the variables are related to each other. The closer the correlation value to 1 or -1, then the strongest relationship between the variables, while the close correlation value to 0, then the weak relationship between the variables. This matrix is also visualized using circle where the size and color of circles depict correlation matrix and the method used to draw this is corrplot() function.



Links Between Costs of Activities and Number of Beds

We also look at the correlation between the Number of Beds and Total Costs. A scatter plot is made with geom_point() function to plot individual points for each hospital against the number of beds for total costs. Linear regression line is also included by using geom_smooth() layer to show the general direction of data points. This enables us to determine whether hospitals with many beds costs more in total or not. This scatter plot is drawn using name colors and labels to ensure that the topic of the plot is well illustrated and visually appealing.

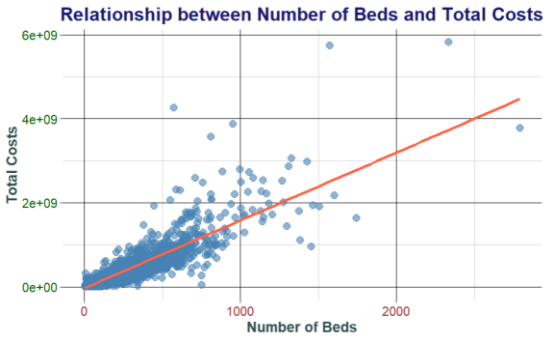


Fig: Relationship between Number of beds and Total Costs

VIII Conclusion

From the analysis of the hospital dataset, there are fundamental insights of the hospitals' financial and operation status. One point of interest is that Total Costs, Total Income, and Net Income appear to share an almost perfect correlation. One important observation made is that most of the larger hospitals are financially endowed hence are likely to earn more revenues based on the realization indicated by the costs ahs presented. This means that the cost of running mainly the large hospitals with many bed capacities are likely to have high costs and at the same time, are also likely to tap more returns on the costs incurred to support the hospitals. Nevertheless, it is established that the increase in costs does not always translate to a similar increase in income hence a need for cost control. Hospitals require operating expenses inorder to minimize their costs while still keenly ensuring they generate

income.

This is a testimony to this idea because while Number of Beds Increase so does the Total Costs because hospitals with large beds need to have the capability of meeting the demands of more patients and offer many services. This finding is consistent with the conventional knowledge that scaling operations in a hospital makes the management more complex and costs higher.

The interpretation of the Total Costs and the Total Income on the scatter plot also confirms general knowledge that the higher the operational cost, the similarly elevated income is not so expected, and this virtually emphasizes the need for efficiency of expenses at the hospital level. Hospitals have to pay attention to such costs since its income growth must accommodate the increased operational costs.

Last, the linear regression model, which tried to forecast Total Costs using only Net Income, shows that in addition to net income, costs increase with the operational scale as well, thus, once again pinpointing the link between the amount of money received and expenditures. They also stress the role of hospitals in working for better cost control, improvement of productivity, and achieving revenue and expenses equality for achieving sustainable financial outcome. Subsequent study on variables such as charity care and patient experience might be instrumental in better understanding administrative solutions to financial management.

IX. Future Work

Some of the areas, which would be interesting for future work, would be the current effect of charity care on the finances of hospitals, specifically identifying its influence on operating expenses and net income. It could also examine other conditionality variables, including patient acuity, clientele and hospital ownership so as to account for appropriateness of financial management practices in varying environments. Synchronous research could explore how such associations change over time and the way in which they are affected by economic or policy shifts. Further, the cross-sectional analysis of hospitals would shed light on how factors such as government support in addition to insurance implications affect financial sustainability, which should benefit policymakers and other administrators working in healthcare.

X. References:

[1] K. Dubas-Jakóbczyk, T. Albreht, D. Behmane, L. Bryndova, A. Dimova, A. Džakula, et al., “The Association Between Hospital Financial Performance and the Quality of Care – A Scoping Literature Review,” Int. J. Health Policy Manag., vol. 11, no. 8, pp. 1284-1295, 2020. doi: 10.34172/ijhpm.2020.51.

[2] J. Gafford, The Impact of Financial Changes on Hospital Quality of Care. Ph.D. dissertation, Univ. of Kansas Medical Center, Kansas City, KS, 2020.

[3] I. C. P. Marques and M. C. Alves, “Hospital Costing Methods: Four Decades of Literature Review,” J. Risk Financial Manag., vol. 16, no. 10, p. 433, 2023. doi: 10.3390/jrfm16100433.

[4] R. Cohen and T. P. O’Brien, “The Impact of the Affordable Care Act on Hospital Efficiency: Evidence from Cost Reports,” Health Serv. Res., vol. 58, no. 5, pp. 879–895, 2023. doi: 10.1111/1475-6773.13899.