JSC370 Midterm

Xiang Gao

2024-03-06

Introduction

Healthcare costs remain a pivotal concern for all stakeholders in the healthcare system, including patients, insurance companies, and governments. Even though standard treatments exist for many common diseases, the costs associated with treating similar patient conditions vary significantly. These variations can be attributed to factors such as the individual clinician's experience, treatment decisions, and broader systemic factors. One major component for guiding treatment policies for critically ill patients is the use of clinical risk scores. Nevertheless, little to no studies exist to investigate the relationship between clinical risk scores and the costs of treatment. This could potentially provide a novel approach for countries with universal healthcare like Canada to more accurately estimate the annual healthcare budget. In addition, it could pose major risks to the general public in countries like the United States seeking healthcare insurance coverage as the insurance company can potentially utilize these risk scores to their advantage for making coverage decisions.

Furthermore, the development and application of clinical risk scores, such as the Model for End-stage Liver Disease (MELD) used in liver transplantation prioritization, demonstrate the profound influence these scores can have on patient outcomes. These scores often serve as proxies of a patient's health status and risk of mortality, with higher scores often reflecting a more severe physiological condition. In this project, we hypothesize that there is a positive association between healthcare costs and the clinical risk scores assigned to patients. Moreover, we aim to examine how this relationship varies across different demographic groups, such as race and gender. Given that many risk scores do not directly include demographic information, the potential disparities and hidden heterogeneities among various subpopulations remain largely unexplored. To address these questions, we will leverage the SUPPORT dataset described below.

The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) was originally a prospective study to develop prognostic models for severely ill hospitalized patients, extending beyond those in intensive care units and emergency departments. Moreover, its impact goes beyond the risk models developed: the dataset collected during the study has served the foundation for studying survival analysis and benchmarking distinct survival algorithms. In our study, instead of analyzing the survival outcomes of patients like many other works, we take a different perspective and look into the hospital charges for each particular patient. The original SUPPORT data was a random sample of 1000 patients collected through the two phases of the study. We use the SUPPORT2 dataset that contains the full 9105 patients. Phase I and II patient data were collected between 1989 and 1994 from multiple institutions including Beth Isreal Hospital in Boston, Duke University Medical Center, and UCLA Medical Center. This dataset provides a unique opportunity to explore the interplay between clinical risk scores and healthcare costs, offering insights that could inform both policy and clinical practice. Details of variables including other response variables like total hospital costs will be discussed further in the subsequent section.

Methods

The raw CSV file of the SUPPORT2 data can be directly downloaded from the link provided by Prof. Frank Harrell - one of the main authors of the study (data acknowledgement: data obtained from http://hbiostat.

org/data courtesy of the Vanderbilt University Department of Biostatistics). Further descriptions of the dataset and corresponding variables can be found through this link. The original SUPPORT research was published on Annals of Internal Medicine and can be found here. We conducted the following analysis in R using RStudio with 16 GB of RAM and Apple M1 Pro; important packages utilized include `ggplot2`, `tidyverse`, `kabelExtra`, and `data.table`.

We load the raw data file using *fread* function from `data.table`. The original data contains 9105 rows and 48 columns (with one extra row ID column, the actual number of covariates is 47), which represents 9105 patients with 47 distinct collected features. For the purpose of this project, we focused on 16 relevant variables detailed below:

- Response Variables:
    1. *charges*: Hospital charges
    2. *totcst*: Total ratio of costs to charges (RCC) cost
    3. *totmcst*: Total micro cost
- Risk Scores:
    4. *sps*: SUPPORT physiology score on day 3
    5. *aps*: APACHE III physiology score on day 3
- Other Information:
    6. *age*: Age of the patients in years
    7. *death*: Death at any time up to December 31 of 1994
    8. *sex*: Gender of the patient
    9. *hospdead*: Death in hospital
    10. *dzgroup*: The patient's disease sub category amogst ARF/MOSF w/Sepsis, CHF, COPD, Cirrhosis, Colon Cancer, Coma, Lung Cancer, MOSF w/Malig
    11. *dzclass*: The patient's disease category amongst "ARF/MOSF", "COPD/CHF/Cirrhosis", "Cancer", "Coma"
    12. *num.co*: The number of simultaneous comorbidities exhibited by the patient
    13. *edu*: Years of education
    14. *income*: Income of the patient in grouped categories
    15. *race*: Race of the patient
    16. *ca*: Whether the patient has cancer (yes), whether it has spread out (metastatic), or if it is healthy (no)

The other variables excluded largely include continuous clinical measurements such as temperature and counts of white blood cells that are already captured by the risk scores. The SPS and APS scores in the dataset both captures the physiological states of patients using day 3 blood test results and other variables. The APS score is developed prior to SPS on intensive care units population whereas SPS applies for more generally hospitalizes patients; the exact equation can be found here.

The data has been imported correctly except for introducing additional ID columns that we removed when subsetting the data. Some missing information in the data is encoded through empty string; therefore, we add it to the list of missing data identifiers to correctly load missing information. We have checked that the data does not contain additional footers or other extra information by examining the top and bottom of the data table. In addition, the headers are correctly identified as we saw from the column names. The response variables, i.e., the costs and the risk scores are continuous numerical variables. Additional information such as race and sex are categorical variables and have been converted to factors in R appropriately.

There are three outcome variables of interest in the SUPPORT2 dataset: hospital charges, total RCC cost, and total micro cost. In the context of the United States, hospital charges refer to the amount billed by the hospital, and they often do not reflect the actual cost of the services and can vary widely between institutions. The RCC cost is an rough estimate of the actual costs of the services calculated from a simple ratio, and

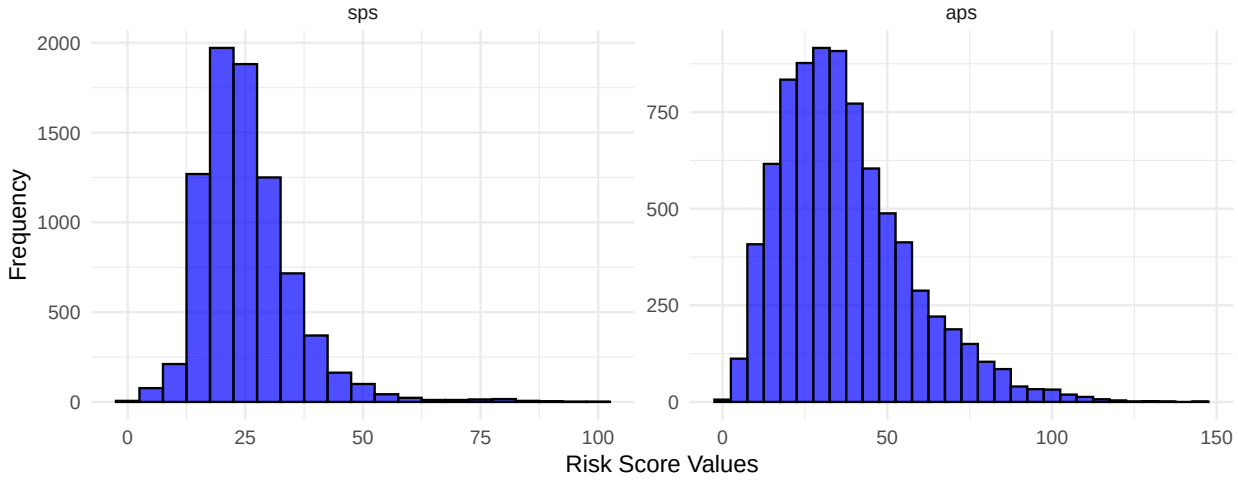
the micro cost is more detailed and precise assessment of the cost of healthcare services by itemizing each every service provided.

Table 1: Summary on 3 Different Cost Variables

Cost Type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	Missing Rate
Hospital Charges	1169.00	9740.000	25024.00	59995.79	64598.00	1435423	172	0.0188907
Total RCC Cost	0.00	5929.566	14452.73	30825.87	36087.94	633212	888	0.0975288
Total Micro Cost	-102.72	5177.404	13223.50	28828.88	34223.60	710682	3475	0.3816584

From Table 1 above, we found that total micro cost has large missing rate of 38%, and it also has invalid values like -102 as costs. Hence, we continue with analyzing charges and total RCC cost and remove the relevant data entries without such information. Additionally, it shows that the billed charges is usually double the amount of costs of services. We will take a closer look of these variables stratified by different subpopulations in the next section.

Figure 1: Empirical Distributions of Two Risk Scores (SPS [left] and APS [right])

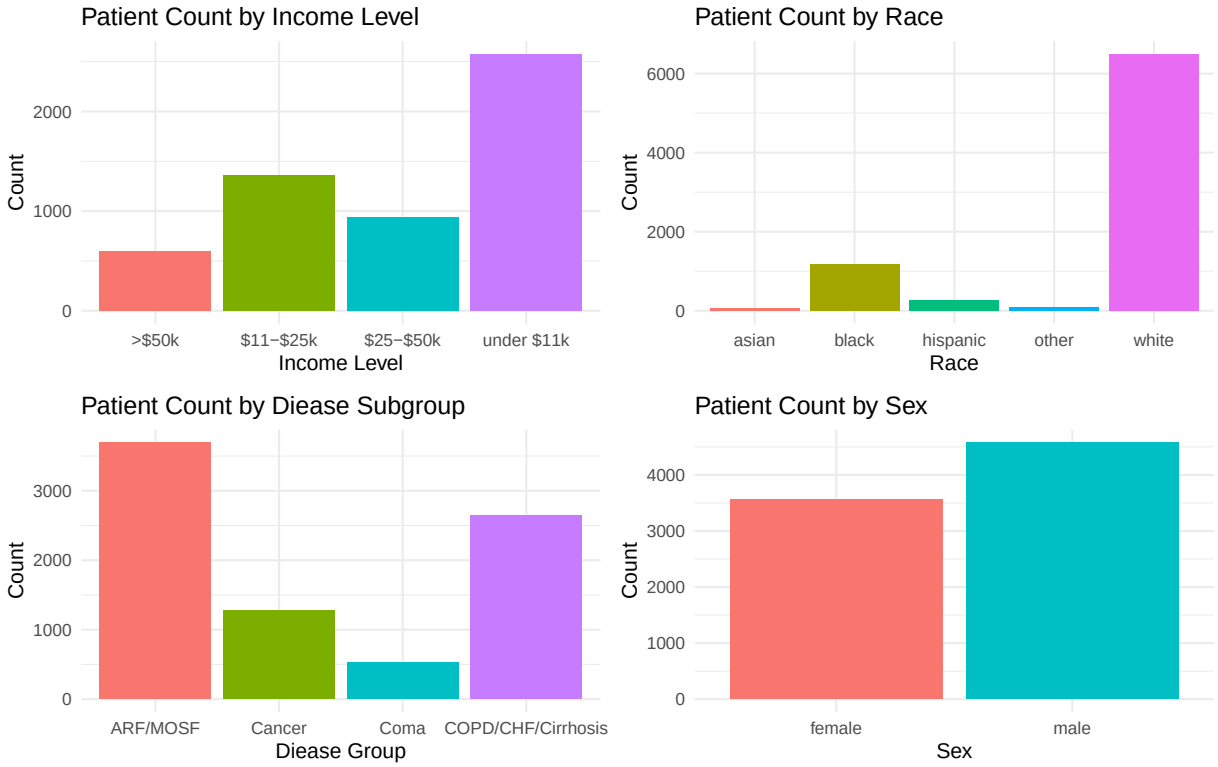


The empirical distribution of the two risk scores are approximately normal skewed to the lower end of the values as shown in Figure 1. After removing the patient records with missing cost information, we observed no missing data for the two risk scores. For other variables of interest like demographics, we have examined the summary of them and found no suspicious values. These variables are either categorical like race and gender or ordinal such as number of years of education and number of comorbidities. Among them, only 3 variables has missing values: *edu* (1472 rows), *income* (2667 rows), and *race* (38 rows). We will defer the treatment of these missing values to a later stage of the project as we gradually obtain a clear picture of what demographic variables demonstrate prominent confounding effect.

Preliminary Results

In this section, we aim to provide a preliminary look at the key variables and their relationships. First, since we have already checked that the empirical distributions for the two risk scores are skewed to the lower end with one single mode, we would like to investigate further whether this changes across race, gender, and income levels. Furthermore, it is crucial to understand if and how these demographic variables influence the hospital costs and charges as it will give us a rough idea of the potential confounders in this study. Nevertheless, we will start with subgroup distributions for some of the selected variables since it may disentangle the true relationship and the sample size of one specific group.

Figure 2: Patient Distribution by Various Subgroups

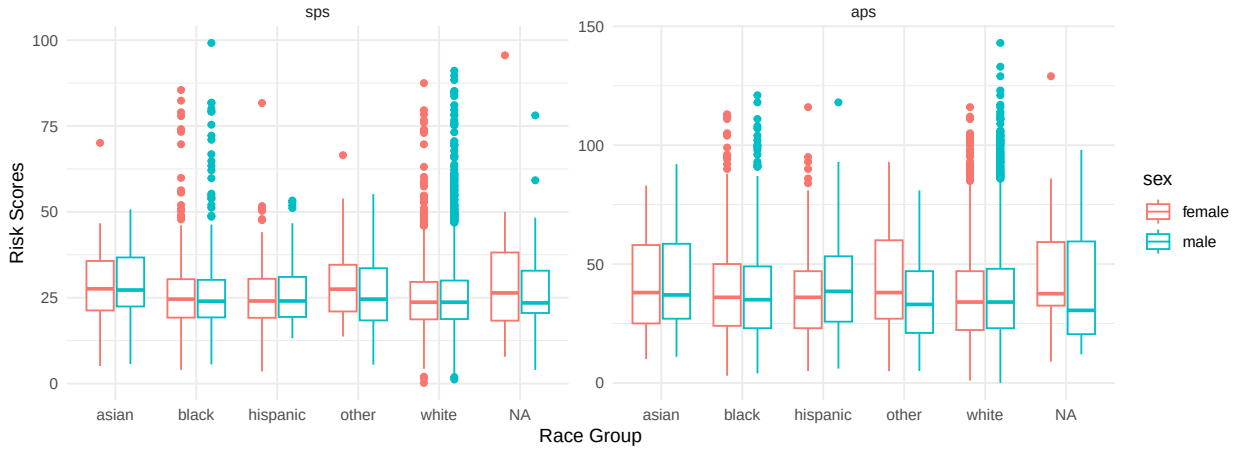


This collection of bar plots provides a breakdown of the patient counts within the SUPPORT2 dataset across various demographic and clinical categories: income level, race, disease subgroup, and sex. Also, I have removed the bars for missing group identity as it will distract the main pattern in some cases.

- **Income**: The majority of patients fall into the lowest income bracket (under $11k). There are considerably fewer patients in the higher income brackets, with the >$50k bracket having the fewest patients. However, with a relatively high missing rate, this might be confounded as higher-income populations refuse to provide their income information.
- **Race**: The dataset predominantly consists of white patients, with the other racial groups represented to a much lesser extent. Asian and Hispanic patients have relatively much smaller representation. The 'other' category could include multiple racial groups not specified in the dataset. The overrepresentation of white patients may affect the generalizability of the study's findings across subgroups.
- **Disease Subgroup**: The bar chart shows that the largest number of patients is grouped under ARF/MOSF (Acute Renal Failure/Multiple Organ System Failure). Chronic conditions like Coma and Cancer subgroups have a similar count, which is less than both ARF/MOSF and COPD/CHF/Cirrhosis (Chronic Obstructive Pulmonary Disease/Congestive Heart Failure/Cirrhosis). The predominance of the ARF/MOSF subgroup reflects the sickness of the hospitalized patients.
- **Sex**: The distribution of patients by sex is relatively balanced, with a slightly higher count of male patients compared to female patients. The near balance in sex distribution is favorable for studying healthcare outcomes without the need for heavy weighting or adjustments.

These charts highlight the importance of considering demographic and clinical context when interpreting data and the potential need for stratification or weighting in analyses to ensure that findings are not biased or confounded by the underlying distribution of the dataset. More importantly, it also raises questions about access to healthcare and socioeconomic factors that may be influencing the makeup of this dataset.

Figure 3: Box Plot of Two Risk Scores by Race and Sex (SPS [left] and APS [right])



The box plot in Figure 3 demonstrates a comparison of the two clinical risk scores, SPS and APS, across different racial groups and between genders. The median scores for both SPS and APS appear to be relatively similar across all racial groups, but the variation for APS is larger as APS has a broader scale from 0-299 (v.s. 0-100). For both scores, there are some outliers, particularly in the 'black' and 'white' categories, indicating some patients with significantly higher risk scores. The interquartile range for APS scores seems to be more consistent between genders within each racial category, suggesting similar variability in APS scores for both males and females.

The SPS and APS scores do not appear to significantly differ across racial groups, which could indicate that these scores are not influenced by race or that the clinical factors these scores represent manifest similarly across these groups. Some out-of-distribution samples indicate that there are patients within each racial category who are at much higher risk than the typical patient. Also, the lack of significant difference between male and female median risk scores may suggest that gender does not play a large role in these particular clinical risk assessments. It is important to note that the variability shown here and in the following figures can be confounded by other variables that indirectly influence both variables of interest and are not taken into account here.

Figure 4: Box Plot of Two Risk Scores by Income Levels (SPS [left] and APS [right])

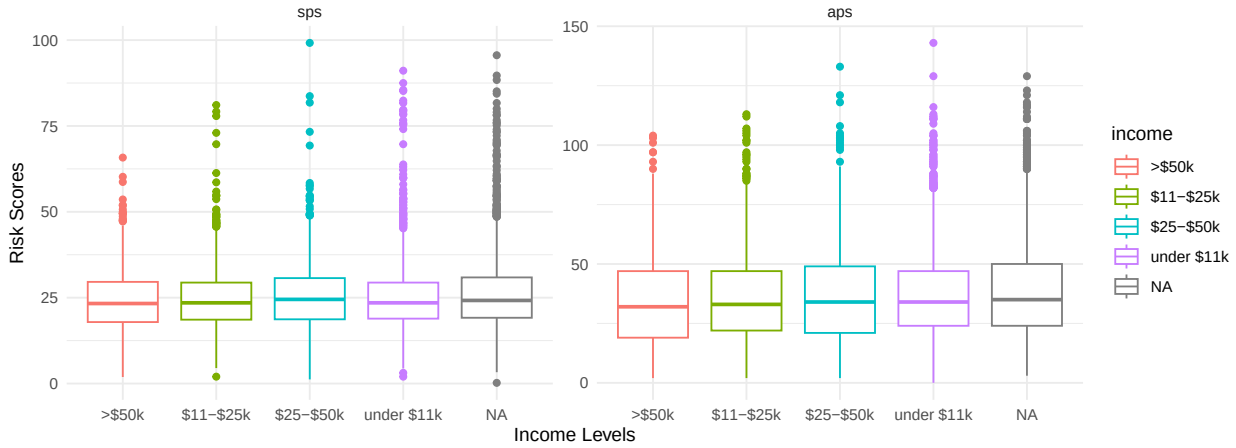
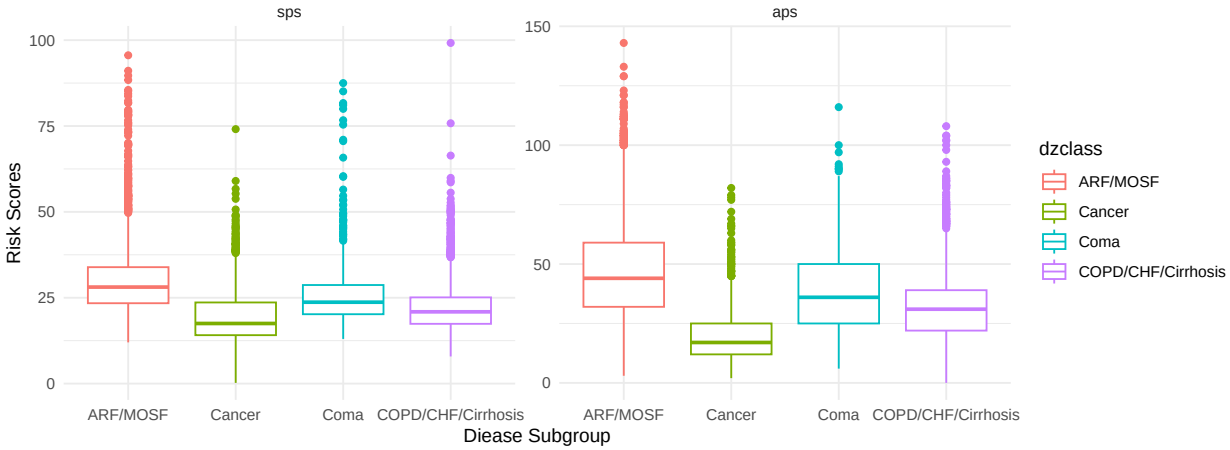


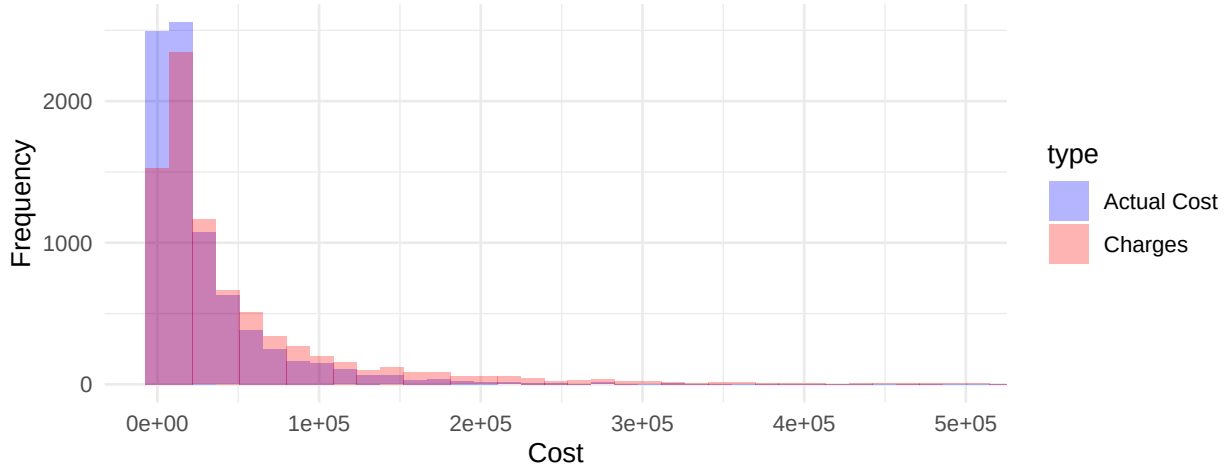
Figure 4 illustrates the distribution of SPS and APS scores across various income levels. There is no immediate visible trend that suggests a significant difference in clinical risk scores based on income levels, although there is a hint of greater variability within the lower income category. The presence of outliers across all income levels suggests that high-risk patients do not belong to any particular income group. However, the number of outliers seems less in the highest income group (> $50k).

Figure 5: Box Plot of Two Risk Scores by Diease Subgroup (SPS [left] and APS [right])



Furthermore, Figure 5 displays the distributions of SPS and APS scores, stratified by disease subgroup: ARF/MOSF (Acute Renal Failure/Multiple Organ System Failure), Cancer, Coma, and COPD/CHF/Cirrhosis (Chronic Obstructive Pulmonary Disease/Congestive Heart Failure/Cirrhosis). Patients with ARF/MOSF have the highest median risk score and the most variability, as evidenced by the length of the box and the range of outliers. There is a clear stratification of clinical risk by disease subgroup, with ARF/MOSF patients typically presenting the highest risk and COPD/CHF/Cirrhosis the lowest, across both scoring systems. This aligns with our intuitions as acute conditions were measured sicker than chronic (i.e. not immediately life-threatening) conditions. The higher number of outliers in the ARF/MOSF and Coma subgroups for SPS may indicate individual patients with complex or acute conditions that significantly increase their clinical risk. It is also critical to consider other clinical and demographic factors that might influence these risk scores.

Figure 6: Charges vs Actual Cost of Each Patient



Now, we turn our attention to the outcome variables. The stacked histogram in Figure 6 compares the empirical distribution of the two outcome variables of interest: hospital charges and actual costs for each patient. The distribution of both charges and actual costs is right-skewed, indicating that a large number of patients incur lower costs, with decreasing frequency as costs increase. There is a visible gap between the hospital charges and the actual costs, with charges consistently higher than the actual costs across all levels. As we mentioned previously, this suggests that what hospitals charge does not directly equate to the actual cost of the services provided. The tail of the charges distribution extends further than that of the actual costs, showing that the highest charges are much greater than the highest actual costs recorded.

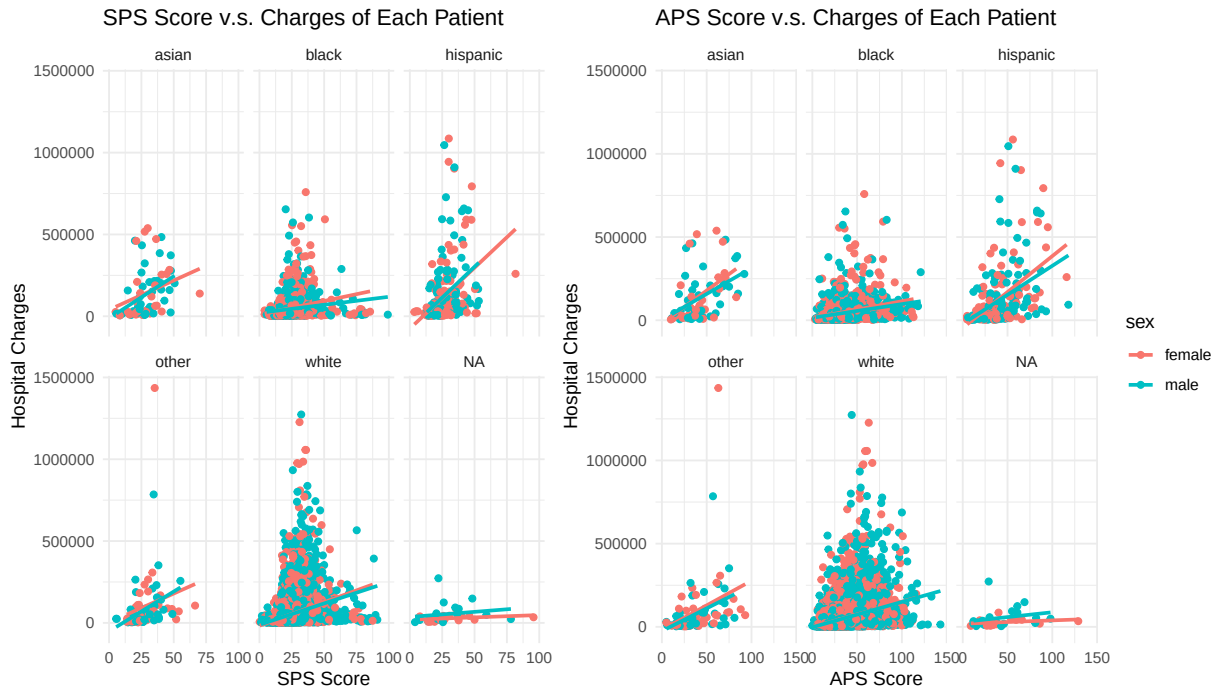
Table 2 provides a summary of hospital charges and costs categorized by racial group. On average, hospital

Table 2: Summary of Hospital Charges and Costs for Each Racial Group

race	Mean Charges	STD Charges	race	Mean Costs	STD Costs
asian	138225.59	150464.47	asian	50930.35	58958.89
black	49913.10	76353.89	black	28492.32	38897.15
hispanic	120295.15	189902.88	hispanic	46895.17	63321.76
other	88280.77	176648.25	other	37188.03	44464.65
white	54385.45	90776.06	white	29814.55	44158.95

charges and costs are the highest for Asian patients within this dataset, followed by Hispanic patients. The standard deviations of charges and costs are also higher for Asian and Hispanic patients, suggesting that these patients experience a broad range of healthcare-related problems and financial burdens. Black patients have the lowest mean charges and costs, which could reflect differences in access to healthcare services, utilization patterns, or underlying health economics factors. There's a substantial difference between the mean charges and mean costs across all racial groups, reflecting the current status quo of the U.S. healthcare billing system. Importantly, additional research would be needed to understand the drivers behind these patterns fully. Factors like insurance coverage, geographic location, type and severity of illness, and socioeconomic status could all play significant roles in these observed differences.

Figure 7: Risk Scores v.s. Hospital Charges by Race and Sex



Based on the scatter plots in Figure 7, the presence of higher charges at higher risk scores which aligns with our intuition suggests that sicker patients, as indicated by their risk scores, tend to incur higher hospital charges. This could reflect more intensive or complex care requirements. Variability in hospital charges at lower risk scores suggests that there could be other determinants of hospital charges besides clinical risk scores, such as the type of treatment, length of stay, or individual hospital billing practices.

In addition, the notable differences between the racial groups in terms of the distribution and slope of the correlation between risk scores and charges may point to underlying disparities in healthcare access and costs. There is also a broad representation of both males and females across the range of charges and risk scores, but distinct patterns based on sex within certain racial categories might indicate gender disparities

Table 3: Linear Regression Result Summary on Charges v.s. SPS Score

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-10490.251	2860.6943	-3.66703	0.0002469
sps	2654.315	105.0838	25.25903	0.0000000

Table 4: Linear Regression Result Summary on Charges v.s. APS Score

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1072.337	2184.9735	-0.4907781	0.6235966
aps	1553.581	51.8255	29.9771515	0.0000000

in healthcare or differences in health status.

We have also observed the same trend between the risk scores and healthcare costs and decided not to include it here. To further validate these interpretations, statistical methods such as regression analysis would be necessary. Additionally, it would be important to control for potential confounding factors including the demographics to isolate the effect of risk scores on hospital charges accurately.

Here, as this is a preliminary result, we analyzed below the relationship between healthcare cost and clinical risk scores with t-tests from linear regression as well as ANOVA F-tests, and we will defer adjusting for demographics to the final report as now we focus on validating and exploring the initial question. We conducted t-tests and F-tests between the two risk scores and hospital charges. We will conduct more analysis on healthcare costs at a later stage, but the conclusion on charges can largely transfer to costs since the costs we used are calculated by applying ratio factors to the charges.

Tables 3 and 4 provide a summary of linear regression results for the relationship between hospital charges and the two different risk scores: SPS and APS. Both SPS and APS scores are significant predictors of hospital charges, as shown by their p-values that are effectively zero. The significant p-values for the risk score coefficients in both models suggest that there is very strong evidence against the null hypothesis (which would typically posit that there is no relationship between the risk scores and charges), leading us to conclude that there is indeed a relationship between the risk scores and hospital charges.

For every one-unit increase in SPS score, the hospital charges are expected to increase by approximately \$2,654.31, holding other factors constant. Similarly, for every unit increase in APS score, hospital charges increase by about \$1,553.58. The negative intercepts in both models indicate that the base charge would be negative, but since risk scores cannot be zero in practice, the negative value for the intercept is a mathematical artifact of the linear model and does not carry a practical interpretation.

Table 5: ANOVA F-test Summary for Charges v.s. SPS

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sps	1	5.527110e+12	5.527110e+12	638.0184	0
Residuals	8141	7.052493e+13	8.662932e+09	NA	NA

Table 6: ANOVA F-test Summary for Charges v.s. APS

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
aps	1	1.524222e+12	1.524222e+12	852.8102	0
Residuals	8141	1.455035e+13	1.787293e+09	NA	NA

The ANOVA F-tests summarized in Tables 5 and 6 showed that for both SPS and APS scores in predicting hospital charges are highly significant, with p-values of 0. This implies that both models are highly unlikely to be due to random chance, confirming the importance of these risk scores in explaining the variability in hospital charges.

These ANOVA results, combined with the linear regression summaries, provide robust statistical evidence of the relationship between both SPS and APS scores and hospital charges. However, it is important to note that while the relationship is statistically significant, this does not imply causation, and other confounding factors could be influencing hospital charges as we saw from the plots.

Summary

Our analysis of the SUPPORT2 dataset has revealed significant associations between clinical risk scores (SPS and APS) and hospital charges, strengthening our hypothesis that higher clinical risk correlates with increased healthcare costs. Across different demographic groups, while risk scores show little variation by income level and gender, there is evidence suggesting that specific disease subgroups — particularly ARF/MOSF may play a role in the variability of charges.

The statistical relationship between risk scores and hospital charges, as evidenced by regression and ANOVA F-tests, indicates that these scores are predictive of the financial burden imposed by hospital stays, independent of patient demographics. For every one-unit increase in SPS and APS scores, hospital charges rise by approximately $2,654 and $1,554, respectively. Additionally, we observed that hospital charges typically exceed the actual costs, with this discrepancy apparent across all racial groups. Asian and Hispanic patients, on average, incur the highest charges and costs, while black patients experience the lowest in both categories, highlighting potential disparities within the healthcare system.

While our statistical findings are significant, they do not imply causation and are subject to the influence of unmeasured confounding factors. Future analyses will aim to adjust for these confounders and explore the causal pathways further. The preliminary evidence points toward the utility of clinical risk scores in predicting hospital charges, which could inform healthcare policy and billing practices. For the next stage of the project, as we have validated the relationship between risk scores and hospital charges, we aim to utilize and develop more advanced models for structured health records, such as Extreme Gradient Boosting. We will also dive deeper into adjusting the relationship with various demographic variables.