The US healthcare sector is one of the largest and fastest growing sectors of the US economy.

The latest available numbers from the National Health Expenditure Accounts (NHEA; the official estimates of total health care spending in the US; see CMS.gov) suggest that this sector contributed $3.8 trillion, i.e. roughly 18% of US GDP, and is growing at about 5% a year. This spending translates to over $11k of expenditure for every American. The following figure shows that healthcare expenditures have grown steadily over the past 20 years.



Hospital care constitutes almost a third of US healthcare expenditures, reaching $1.2 trillion in 2019.

A large and potentially avoidable contributor to hospital expenses is readmissions, estimated to be $26 billion annually (Wilson, 2019). A hospital readmission occurs when a patient is admitted to a hospital within a specified time period after being discharged from an earlier (initial) hospitalization. For Medicare, this time period is defined as 30 days. Medicare uses an “all-cause” definition of readmission, meaning that hospital stays within 30 days of a discharge from an initial hospitalization are considered readmissions, regardless of the reason for the readmission. For the purposes of this study we follow the same definition.

Hospital readmissions are particularly acute among seniors and those suffering from certain ailments such as Pneumonia and Myocardial Infarction. These readmissions put significant pressure on the Medicare system, which is the payee for most of these costs. Under the Affordable Care Act, the Hospital Readmissions Reduction Program was instituted to rein in readmission costs. The program evaluates hospitals on readmission rates and penalizes those which do not meet certain pre-specified criteria.

It is not surprising that in order to curtail costs and avoid fines, hospitals are trying to understand the factors driving readmissions of patients. One of the largest areas for improvement exists in patients suffering from chronic conditions such as asthma, arthritis, and diabetes.

We will use a publicly available dataset to understand why some patients are more likely to be readmitted than others. The dataset represents 10 years (1999-2008) of clinical care of 68.654 patients at 130 US hospitals and integrated delivery networks. There are a total of 97,650 observations in the data. In this instance we are only examining patients suffering from diabetes and related complications. The data contains such attributes as patient and encounter number, whether readmitted subsequently after the current visit discharge, demographics of patients, past medical history of patient, and information on HA1C test and medicine compliance collected during the current admission. The dataset is a subset of one submitted on behalf of the Center for Clinical and Translational Research, Virginia Commonwealth University, a recipient of NIH CTSA grant UL1 TR00058 and a recipient of the CERNER data.

[https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008#](https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008)

Detailed definitions of variables in the dataset are available on the next page. We will model these data to understand which factors can explain variation in readmissions rates amongst these patients.

1. Using simple data analyses techniques, such as t-tests and cross-classification tables (hint: use the Fit Y-X feature in JMP for both tasks), answer the following questions:
   1. What percentage of African American patients are readmitted?
   2. Which race(s) have a higher rate of representation among readmitted patients than in the overall population of patients?
   3. What percentage of the total sample consists of patients who are listed as other/unknown gender and are not readmitted?
   4. What is the mean number of diagnoses among patients who have been readmitted?
   5. Perform a t-test in JMP to determine if the time spent in hospital for those readmitted is statistically different from those who did not get readmitted.
   6. What percentage of readmitted patients had normal HbA1C and steady medicine compliance?
2. Run a logistic regression of the Readmitted variable against all available independent variables. You will need to create indicator variables representing race and gender information. Using the logistic regression output:
   * 1. Is the model significant? Please justify your answer with number(s).
     2. How good is the model fit? Please justify your answer with number(s).
     3. Is the coefficient for the Hispanic indicator statistically significant? Please justify your answer with number(s).
3. Below is the output from a logistic regression model with Readmitted as the dependent variable and all other variables included as potential independent variables. The logistic regression was refined using backwards elimination based on p-values with a 5% significance threshold. Answer the following questions based on this output.
   1. Interpret the coefficient for Num\_Procedures. Translate the odds ratio and explain in simple English what it means.
   2. What do the coefficient and odds ratio for the indicator variable Normal\_Steady tell us? Interpret in plain English. Make sure to carefully read the variable description of the variable Normal\_Steady.



1. Provide specific recommendations to hospital administration about how to reduce hospital readmissions. Your response should be less than 150 words and should reference specific analysis based on the data.
2. **Data Description**

|  |  |
| --- | --- |
| **Variable**  **Name** | **Description** |
| **Unique Patient Identifier** | |
| Patient\_Id | Unique patient identification number |
| Encounter\_Id | Each hospital admission’s unique identification number |
| **Dependent Variable** | |
| Readmitted | Whether the patient was readmitted within 30 days of this discharge to an inpatient hospital facility (1=Yes, 0=No) |
| **Patient Demographics** | |
| Race | Race of the patient (African American, Asian, Caucasian, Hispanic, and Other) |
| Gender | Gender of the patient (Male, Female, Other/Unknown) |
| Age | Age group of the patient (1=0-10; 2=10-20; 3=20-30; 4=30-40; 5=40-50; 6=50-60; 7=60-70; 8=70-80; 9=80-90; 10=90-100) |
| **Patient History** | |
| Time\_Hospital | Number of days spent in hospital in previous 12 months window |
| Num\_Procedures | Number of medical procedures performed on the patient in previous 12 months window |
| Number\_Diagnoses | Number of medical diagnoses determined on the patient in previous 12 months window |
| Number\_Outpatient | Number of outpatient visits made by the patient to a hospital in previous 12 months |
| Number\_Emergency | Number of emergency room visits made by the patient in previous 12 months |
| Number\_Inpatient | Number of inpatient visits made by the patient to a hospital in previous 12 months |
| **Visit Details** | |
| High\_Steady | Indicator suggesting whether patient had high HbA1C and whether compliance with diabetes medication was steady, No HbA1C test performed is base case (see note below) |
| High\_Down | Indicator suggesting whether patient had high HbA1C and whether compliance with diabetes medication was down, No HbA1C test performed is base case (see note below) |
| Normal\_Steady | Indicator suggesting whether patient had normal HbA1C and whether compliance with diabetes medication was steady, No HbA1C test performed is base case (see note below) |
| Normal\_Down | Indicator suggesting whether patient had normal HbA1C and whether compliance with diabetes medication was down, No HbA1C test performed is base case (see note below) |

1. Note: Information on a patient’s HbA1C and diabetes medication compliance is only available if a HbA1C test was performed. Note that this set of indicator variables represents five categories: (1) Test revealed high HbA1C and steady compliance with medication, (2) Test revealed high HbA1C and down compliance, (3) Test revealed normal HbA1C and steady compliance, (4) Test revealed normal HbA1C and down compliance, and (5) No HbA1C test was performed.