D205 Data Acquisition

Performance Assessment

Robert C. Miller

Western Governors University

Research Question

The research question I wanted to answer for this thesis is, "What is the total number of patients readmitted and out of that group how many where overweight, Diabetic, had Anxiety, Asthma, High Blood Pressure or a Stroke?"

This will be an analysis of whether a patient was readmitted and what contributing factors were present. This is a truly relevant question for both the hospital as a business and the Insurance Providers. According to a report by Cureatr, Medicare readmissions totaled 2.3 million, with an average cost of $15,500 per readmission. Medicaid accounted for 721,399 readmissions, with an average cost of $14,100. Using this data, physicians could make recommendations to patients after discharge to address these risk factors. It can also provide an "order of importance" to be addressed upon discharge. The most common factors can be addressed first. There is also a financial benefit to the hospital. As a part of the Affordable Care Act, Medicare and Medicaid have begun to reduce payments to hospitals with excessive rehospitalization rates (Cureatr, n.d.).

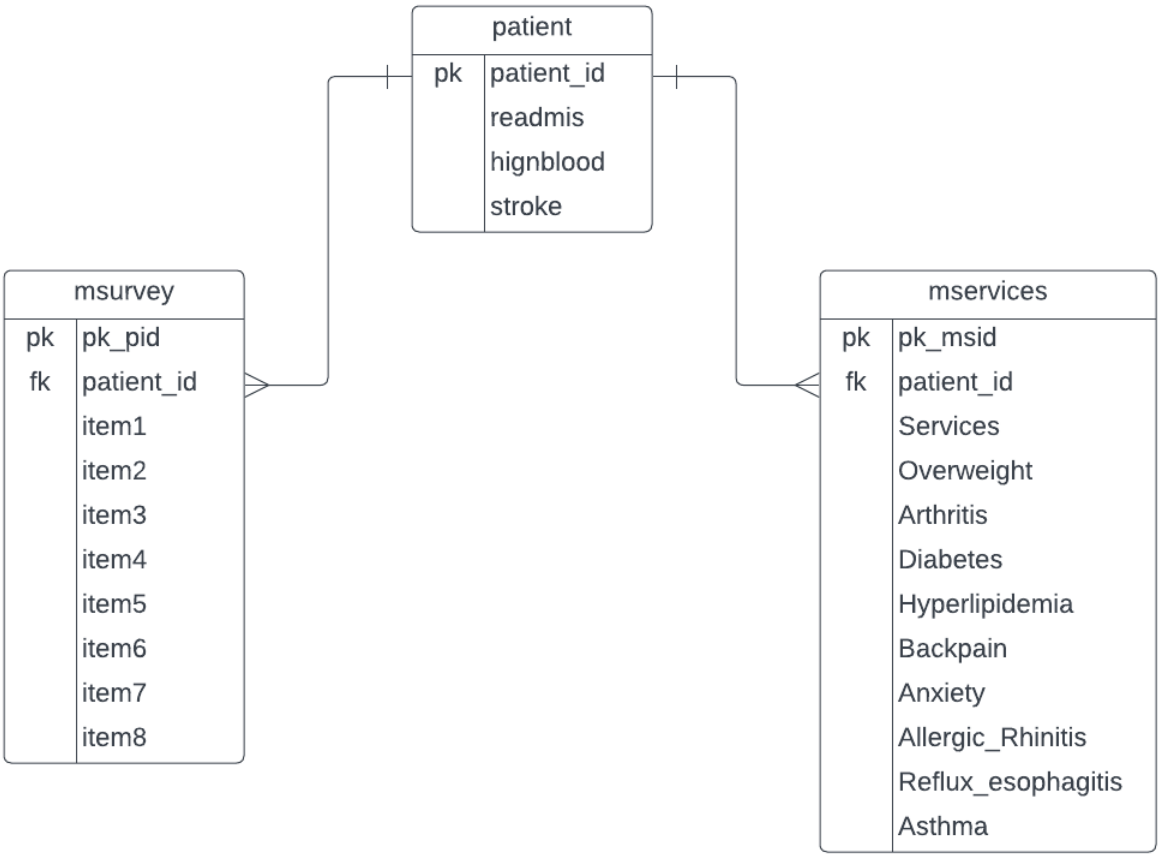
A: Identifying Data

This research will require a few columns from the available data. The columns used from the patient table will be patient_id, readmis, hignblood and stroke each with a data type of text. From the mservices table we will be using patient_id, overweight, arthritis, diabetes, anxiety, and asthma. Patient_id is the primary key for the patient table. We will be using that as the foreign key for both the mservices and msurvey tables to keep integrity of the data between these tables. Both mservices and msurvey will have their own unique primary keys while maintaining Patient_ID as the foreign key. We will be importing both the msurvey.csv and mservices.csv files into the existing database although we will not need the msurvey data to answer the current question, it may be useful later. Each of these columns will be imported as varchar for simplicities sake. We will first get a total count of the total number of readmits. In that query we will also include conditional statements that will count the number of patients where readmission was true as well as another factor. These will give us our Data to work from, the total number of readmissions and the totals of each relevant condition that the patient presented with at that time.

B. Logical Data Model

The CSV file mservices.csv is the primary one that will be used to answer my chosen SQL question. I will be importing the entire file and using all columns from this table. We will also be using the

patient table from the existing database and joining against the patient_id as that is a unique identifier in both tables.



B1: SQL Code for Table Creation

```
CREATE TABLE msurvey(,
    patient_id varchar,
    item1 varchar,
    item2 varchar,
    item3 varchar,
    item4 varchar,
    item5 varchar,
    item6 varchar,
    item7 varchar,
    item8 varchar,
    FOREIGN KEY (patient_id) References public.patient(patient_id)
);
```

```
CREATE TABLE mservices(
    patient_id varchar,
    Services varchar,
    Overweight varchar,
    Arthritis varchar,
    Diabetes varchar,
    Hyperlipidemia varchar,
    Backpain varchar,
    Anxiety varchar,
    Allergic_Rhinitis varchar,
    Reflux_esophagitis varchar,
    Asthma varchar,
    FOREIGN KEY (patient_id) REFERENCES public.patient(patient_id)
);
```

B2: Loading the CSV Data

```
--importing both the msurvey.csv file and mservices.csv file into the database
Copy msurvey
FROM 'C:\LabFiles\Medical\msurvey.csv'
DELIMITER ','
CSV HEADER;

COPY mservices
FROM 'C:\LabFiles\Medical\mservices.csv'
DELIMITER ','
CSV HEADER;

--After import add an auto-incrementing primary key to each table
ALTER TABLE msurvey
ADD COLUMN pk_pid SERIAL PRIMARY KEY;

ALTER TABLE mservices
ADD COLUMN pk_msid SERIAL PRIMARY KEY;
```

C. SQL Query

```
SELECT
    COUNT(p1.readmis) AS TotalReadmits,
    COUNT(CASE WHEN m1.overweight = 'Yes' THEN 1 END) AS OverweightReadmits,
    COUNT(CASE WHEN m1.Diabetes = 'Yes' THEN 1 END) AS DiabetesReadmits,
    COUNT(CASE WHEN m1.Anxiety = 'Yes' THEN 1 END) AS AnxietyReadmits,
    COUNT(CASE WHEN m1.Asthma = 'Yes' THEN 1 END) AS AsthmaReadmits,
    COUNT(CASE WHEN p1.hignblood = 'Yes' THEN 1 END) AS HBPReadmits,
    COUNT(CASE WHEN p1.stroke = 'Yes' THEN 1 END) AS StrokeReadmits
FROM
    patient AS p1
FULL JOIN
    mservices AS m1 ON p1.patient_id = m1.patient_id
WHERE
    p1.readmis = 'Yes';
```

The SQL statements above help us to answer our initial SQL question of: "What is the total number of patients readmitted and out of that group how many where overweight, Diabetic, had Anxiety, Asthma, High Blood Pressure or a Stroke"

The patient_id is the primary key in the patient table from the original database, we will be pairing that with the mservices data by joining against that patientID since it is a unique identifier. In order to answer our stated question, we are only looking at records where readmission is equal to a yes. Among those we will be looking at some of the comorbidities presented in the data and how many of the readmitted patients also present with these conditions.

C1: CSV FILE

Please see the included .csv file for the output of this query.

D: Data Refresh Interval

We are looking for the total count of readmitted patients and then among that population how many presented with each of a few comorbidities. Using this data, we can provide a physician a hierarchy of diagnoses. These numbers can give them a list to go by of what issues are present in their population and their frequency among this population of patients who are often hospitalized. Patient data is also not likely to change very quickly since health, and health care, require long term monitoring. We would want to update the Medical_View table quarterly and rerun this report at that time to recommend any novel changes. The mservices and msurvey CSV files can be updated daily; these would be working files that can be kept live by

the staff. These flat files can be re-imported, and the data can be cleaned to remove non-essential duplications of data. For the current research question, we would only need to update the results on a quarterly basis; these updated reports could be used by the physicians to see what comorbidities are prevalent in their communities so that they can be addressed, potentially prior to admission. This could help reduce the number of patients seen for these comorbidities in total.

E. Panopto Video

Please see the included video

F. none used

G. Resources

- Cureatr. (June 2023). Cost of Hospital Readmissions: What the Statistics Tell Us. *Cureatr Blog*. https://blog.cureatr.com/cost-of-hospital-readmissions-what-the-statistics-tell-us
- Marshall MN, Shekelle PG, Brook RH, Leatherman S. Use of Performance Data to Change Physician Behavior—Reply. JAMA. 2000;284(9):1079. doi:10.1001/jama.284.9.1079
- Warchol, S. J., Monestime, J. P., Mayer, R. W., & Chien, W. W. (2019). Strategies to Reduce Hospital Readmission Rates in a Non-Medicaid-Expansion State. Perspectives in health information management, 16(Summer), 1a.