

# Data Acquisition Performance Evaluation

D205

TRESA (TESSIE) AUSTIN

WGU

January 31, 2024

A. SQL research question

This research question will be derived from medical data files, patient table in PG Admin and the add on file mservices of medical services information. This research question will answer the question using a SQL select statement joining two tables from the aforementioned files. The research question "How many overweight male patients had an initial hospital stay of more than 30 days?" will be the focus of this project.

A1. Data used for research question

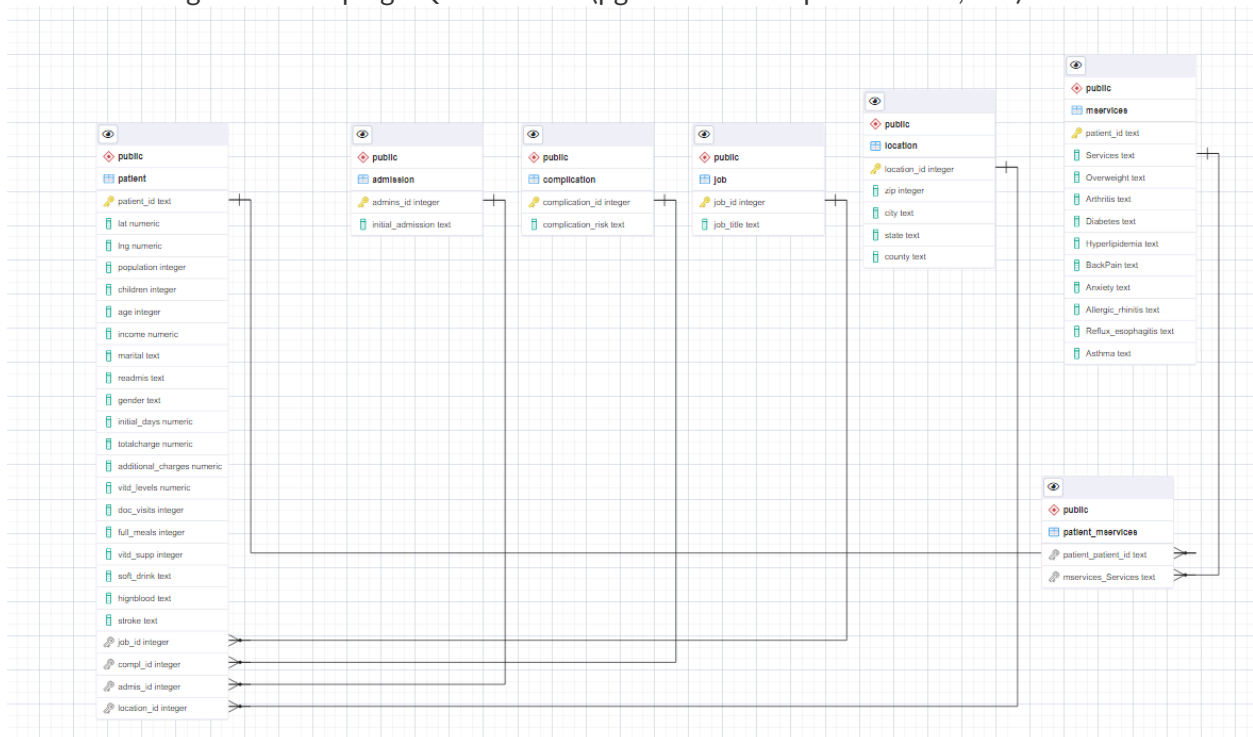
The tables that will be used are the patient table and admission table from medical\_data and mservices created from the add on CSV file mservices.csv.

Column from patient table used are patient\_id, gender, and initial\_days and columns from mservices table used are patient\_id, and overweight will be joined. The following variables patient\_ID, gender, and overweight are of data type text as identified in the Entity Relationship Diagram below in section B whereas initial\_days is data type numeric.

B. Entity relationship diagram (ERD)

To create an entity relationship diagram, data from PG Admin as well as data from additional lab files were used. Entity relationship diagrams help illustrate how data points relate to each other and help define the code that needs to be written ("ER diagrams," n.d.).

PG Admin's built in ERD tool can provide a graphical representation of tables to help in building and developing SQL databases (pgAdmin Development Team, n.d.)



B1. SQL code to create table based on ERD

The SQL code needed to create the table based on the ERD above is listed below. The data lists each column and relevant primary key. Also listed is whether the field can be null or not.

```
CREATE TABLE public.mservices
(
    patient_id text COLLATE pg_catalog."default" NOT NULL,
    "Services" text COLLATE pg_catalog."default",
    "Overweight" text COLLATE pg_catalog."default",
    "Arthritis" text COLLATE pg_catalog."default",
    "Diabetes" text COLLATE pg_catalog."default",
    "Hyperlipidemia" text COLLATE pg_catalog."default",
    "BackPain" text COLLATE pg_catalog."default",
    "Anxiety" text COLLATE pg_catalog."default",
    "Allergic_rhinitis" text COLLATE pg_catalog."default",
    "Reflux_esophagitis" text COLLATE pg_catalog."default",
    "Asthma" text COLLATE pg_catalog."default",
    CONSTRAINT mservices_pkey PRIMARY KEY (patient_id)
)

TABLESPACE pg_default;

ALTER TABLE public.mservices
    OWNER to postgres;
```

B2. SQL code to load data to table created in B1

The SQL code needed to load data from the add-on CSV table used, mservices, is listed below. This code loads all records into the table keeping referential integrity.

```
copy public.mservices (patient_id, \"Services\", \"Overweight\", \"Arthritis\",
\"Diabetes\", \"Hyperlipidemia\", \"BackPain\", \"Anxiety\", \"Allergic_rhinitis\",
\"Reflux_esophagitis\", \"Asthma\") FROM 'C:/LabFiles/Medical/MSERVI~1.CSV' CSV
QUOTE \"\" ESCAPE \"\", \"\"
```

C. SQL query to obtain answer to SQL research question

To answer the SQL research question outlined in part A. SQL code needs to be derived to answer, “How many overweight male patients had an initial hospital stay of more than 30

days?" To produce an answer, the following SQL code is needed:

```
SELECT
    ms."Overweight",
    pat.patient_id,
    pat.gender,
    pat.initial_days
FROM patient AS pat
JOIN mservices AS ms ON pat.patient_id = ms.patient_id
WHERE pat.initial_days >= 30
AND pat.gender = 'Male'
```

- C1. A CSV file of the data from the above query has been submitted as requested. This CSV file contains the information that answers the research question based on the SQL code above. The data shows that 2415 male patients had a hospital stay longer than 30 days which is approximately 25% of the patients in the study.

- D. Time period to refresh data

As this data can be used to understand co-morbidity factors for men as it pertains to lengthy hospital stays. Data would best be refreshed at least quarterly.

- D1. Time period best for refreshing data

Refreshing the data quarterly would allow medical professionals to see trends in men's health and determine if a pattern emerges of overweight men and hospital stays longer than 30 days.

- E. Panopto video

Video link of execution of SQL code to produce the results for "How many overweight male patients had an initial hospital stay of more than 30 days?" is provided below.

- E1. Video Link

Below is the link to a video demonstration of creating the table, loading data in the table and executing the query to generate results for the SQL research question.

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=a948d360-0014-4496-a2bd-b116017bd29c>

- F. No additional web sources were used to acquire data.

- G. Sources

Lucidchart. (n.d.). ER diagrams. Lucidchart. <https://www.lucidchart.com/pages/er-diagrams>

pgAdmin Development Team. (n.d.). ERD tool. pgAdmin. Retrieved from [https://www.pgadmin.org/docs/pgadmin4/development/erd\\_tool.html](https://www.pgadmin.org/docs/pgadmin4/development/erd_tool.html)