D205 Performance Assessment

Data Acquisition

Instructor: Dr. Sewell

Student Name: Nicole Haibach

Student ID: 001260374

Email: nhaibac@wgu.edu

1. **Research Question**

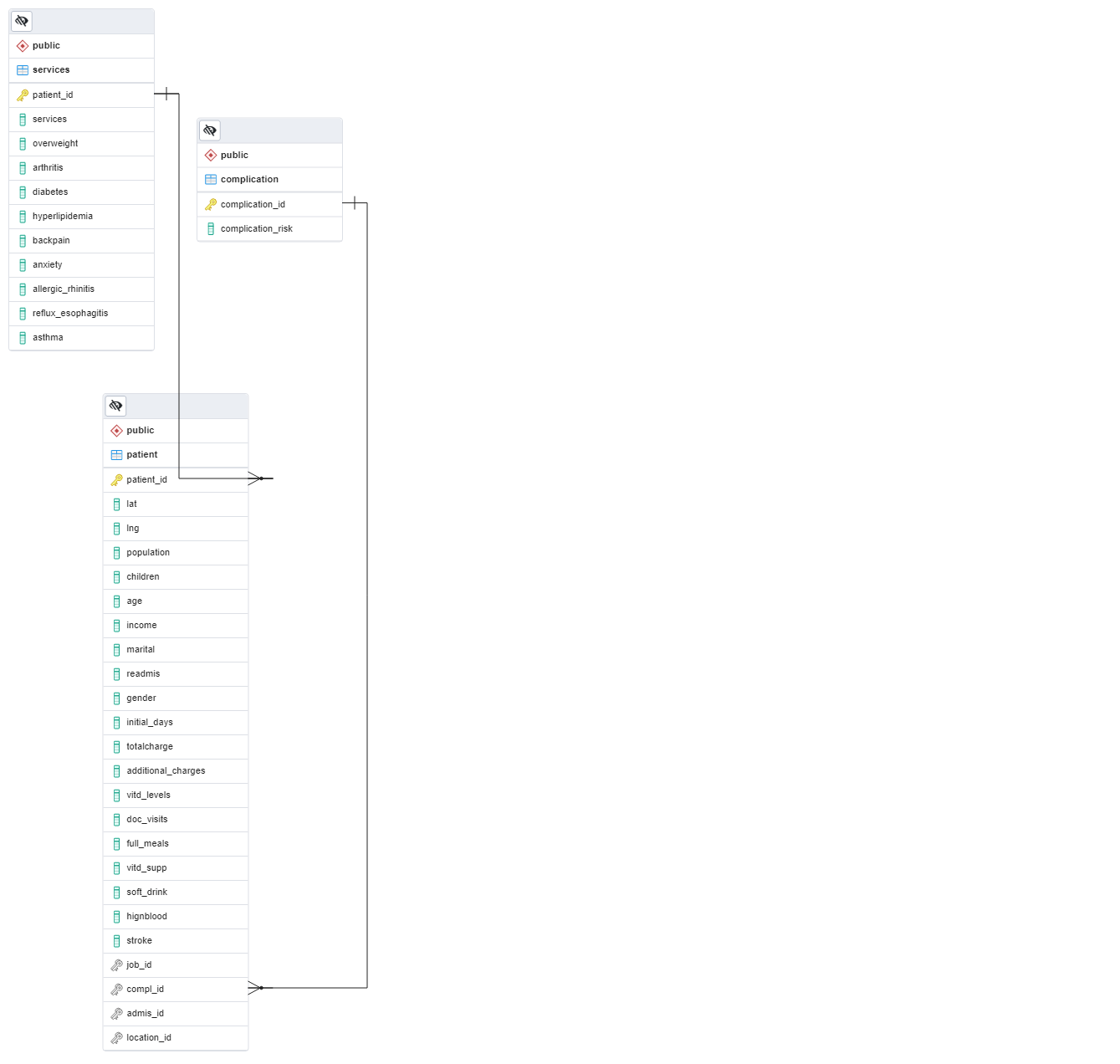
My research question for this assessment will be evaluating services based on high complication risk. What medical service will most likely be needed by the patient based on their complication risk being labeled as high? I will look at what primary medical service is the most common for patients who have a high-risk complication status. This will pull data from the medical services table and the customer information table both of which come from the Western Governors University data “Medical Readmission.”

**A1- Identifying Data:**

The research question will compare the services and patient information tables based on customer ID. The data that will be pulled from the original data set of patient information will be the complication ID which is an integer data type. It will be set to only include the complication ID equal to two, which is the number for high-risk status. The information on complication status is in the complication table that lists two being equivalent to high-risk status. The CSV file used was the mservices file which contained information on primary medical service performed. The first column that was used is the patient ID column in both the CSV file and the patient table is varchar with letters and numbers to have a unique identifier for each patient. The second column used from the CSV file was text of varying lengths to denote the services, which is why varchar was used. The CSV file was imported as a new table labeled services where the services column was used. The comparison of the tables was accomplished by joining the tables with the use of the patient ID columns of both tables.

1. **Logical Data Model:**

The Entity-Relationship Diagram (ERD) tool is a database design tool that provides a graphical representation of database tables, columns, and inter-relationships. ERD can give sufficient information for the database administrator to follow when developing and maintaining the database (pgAdmin 4 Documentation 2023). The tables used were the patient, services, and complication tables. The complication table was just used as a visual reference for what the complication IDs meant. The patient ID and complication ID columns of the patient table were used to link to the services table and to identify which patients were of high risk. The services and patient ID columns of the services table were used to link the patients to the patient table and to identify which primary services were completed.



**B1- Code for the physical data model:**

The code below created an ERD of the connections needed to answer my research question.

CREATE TABLE public.complication

(

complication\_id integer NOT NULL,

complication\_risk text,

PRIMARY KEY (complication\_id)

);

CREATE TABLE public.patient

(

patient\_id text NOT NULL,

lat numeric,

lng numeric,

population integer,

children integer,

age integer,

income numeric,

marital text,

readmis text,

gender text,

initial\_days numeric,

totalcharge numeric,

additional\_charges numeric,

vitd\_levels numeric,

doc\_visits integer,

full\_meals integer,

vitd\_supp integer,

soft\_drink text,

hignblood text,

stroke text,

job\_id integer,

compl\_id integer,

admis\_id integer,

location\_id integer,

PRIMARY KEY (patient\_id)

);

CREATE TABLE public.services

(

patient\_id character varying NOT NULL,

services character varying,

overweight character varying,

arthritis character varying,

diabetes character varying,

hyperlipidemia character varying,

backpain character varying,

anxiety character varying,

allergic\_rhinitis character varying,

reflux\_esophagitis character varying,

asthma character varying,

PRIMARY KEY (patient\_id)

);

ALTER TABLE public.services

ADD FOREIGN KEY (patient\_id)

REFERENCES public.patient (patient\_id)

NOT VALID;

ALTER TABLE public.patient

ADD FOREIGN KEY (compl\_id)

REFERENCES public.complication (complication\_id)

NOT VALID;

ALTER TABLE public.patient

ADD FOREIGN KEY (patient\_id)

REFERENCES public.services (patient\_id)

NOT VALID;

END;

**B2- Loading CSV data**

The code denotes what needs to be done to the CSV file to upload it into a table to be usable.

-- Table: public.services

DROP TABLE public.services;

CREATE TABLE public.services

(

patient\_id varchar(50) NOT NULL,

services varchar(50),

overweight varchar(3),

arthritis varchar(3),

diabetes varchar(3),

hyperlipidemia varchar(3),

backpain varchar(3),

anxiety varchar(3),

allergic\_rhinitis varchar(3),

reflux\_esophagitis varchar(3),

asthma varchar(3),

CONSTRAINT services\_pkey PRIMARY KEY (patient\_id)

)

TABLESPACE pg\_default;

ALTER TABLE public.services

ADD FOREIGN KEY (patient\_id)

REFERENCES public.patient (patient\_id)

NOT VALID;

ALTER TABLE public.services

OWNER to postgres;

--command " "\\copy public.services (patient\_id, services, overweight, arthritis, diabetes, hyperlipidemia, backpain, anxiety, allergic\_rhinitis, reflux\_esophagitis, asthma) FROM 'C:/Users/LabUser/Desktop/MSERVI~1.CSV' DELIMITER ',' CSV HEADER QUOTE '\"' ESCAPE '''';""

1. **Answering Query**

The question that was being answered was what are the primary services being given to patients who are considered high risk according to the complication ID. This was accomplished by selecting for services and counting all services. Joining the patient table on the patient ID to compare the patient and services table. Then add in the where clause that the complication ID is equal to two, which is the high-risk category. The results were grouped by services to separate the services of the high-risk category. Finally, the results were put in descending order to display the primary service that is conducted for high-risk patients.

SELECT s.services, COUNT(services)

FROM services AS s

JOIN patient as p

ON p.patient\_id = s.patient\_id

where compl\_id = 2

GROUP BY s.services

ORDER BY COUNT DESC;

**C1- Results of Query**

The results of the query displayed that there are four primary services that high-risk patients receive. The most common service is blood work with 1105 total patients. The next services are intravenous, CT scans, and MRI respectively. They can use these numbers to help decide what supplies they might need to focus on for their high-risk patients.

1. **Time period:**

The add-on file should be refreshed on a biannual basis to keep the data relevant for a check every six months to see how the high-risk patients are taken care of over the period.

**D1- Time Period Explanation-**

The numbers should be run on a biannual basis to evaluate the number of high-risk patients and what treatments need to be ready. Being prepared for what your high-risk patients need will help the hospital overall keep a good survival rate. The high-risk patients are going to be the patients that need the most help and supplies. If the hospital can stay on top of what the patients typically need, they will be able to keep those supplies available.

**G and H- References**

Western Governors University. (2015). *Medical Readmission* [data set]. https://access.wgu.edu/ASP3/aap/content/g9rke9s0rlc9ejd92md0.html

*pgAdmin 4 Documentation*. (2023, September 20). Retrieved from https://www.pgadmin.org/docs/pgadmin4/development/erd\_tool.html