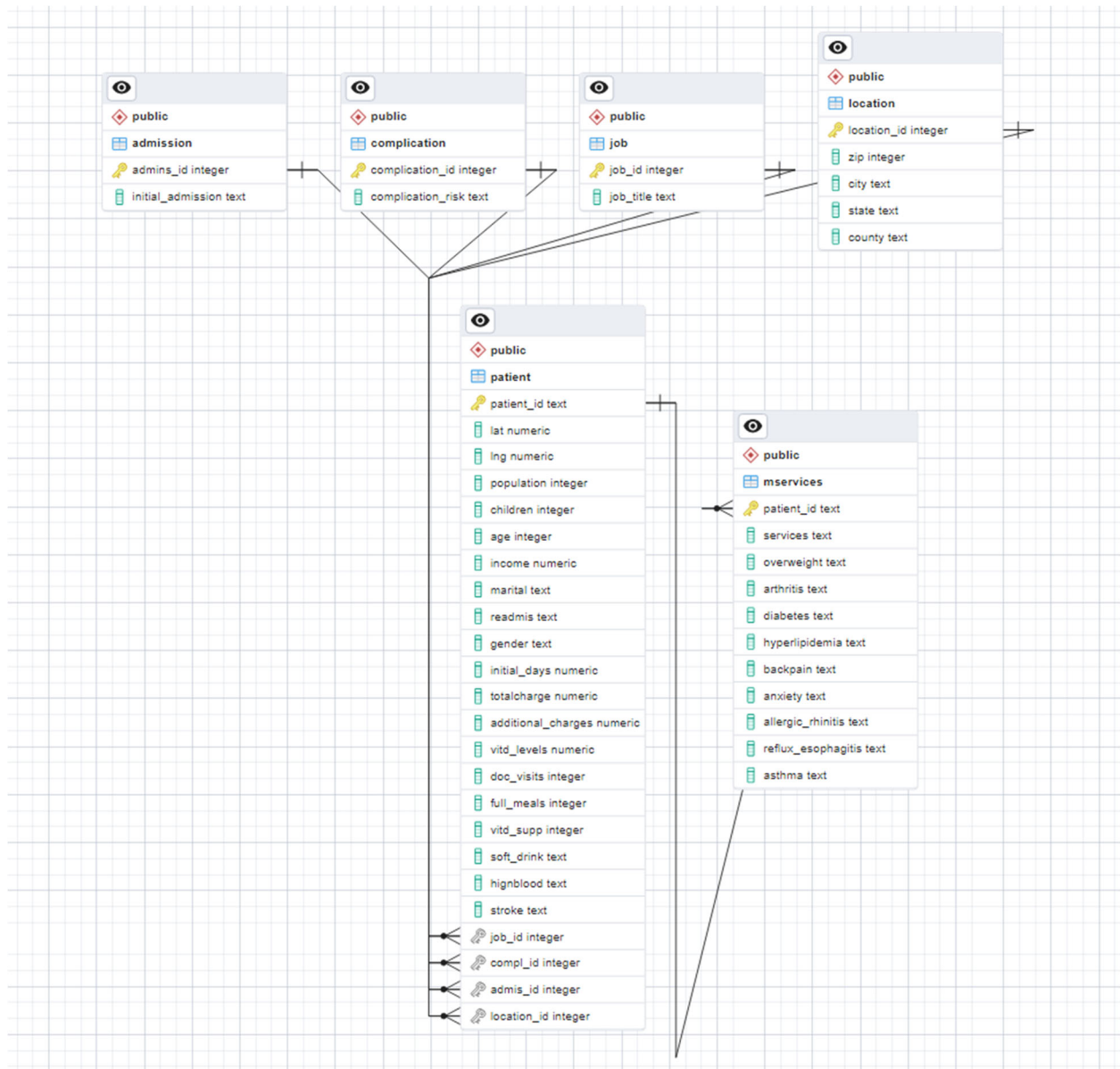


A. Question: How does the prevalence of specific health conditions (like Diabetes or Arthritis) vary across different age groups and income brackets within the patient population?

A1. To answer this question, I need info from both the original medical_data database and the external mservices.csv file. In the database, I'll investigate the patient table, specifically looking at the patient_id, age, and income. The patient_id is the unique id for each patient, which is important for identifying data from both the patient table and mservices table. The age and income columns offer demographic and economic context for the patients. In the mservices.csv file, the patient_id is important again for establishing linkage with the patient table, and the inclusion of the diabetes and arthritis columns allows me to pinpoint individuals affected by these conditions.

B. The ERD involves the patient table from the medical_data database and the new table created from the mservices.csv file. The patient_id serves as the foreign key in the mservices table linking back to the patient table. This establishes a one-to-many relationship (one patient can have multiple service records).



B1. The SQL below creates a new table with headers that correspond with the mservices.csv file. The first part of the query ensures that a duplicate table is not created by creating the table only if it already doesn't exist. The primary key for this table will be patient_id and every column will be of type text including the patient_id.

```
CREATE TABLE IF NOT EXISTS 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)
)
```

B2. This SQL code loads the mservices.csv file into the newly created table and ensure the primary key is patient_id. I used the GUI in pgadmin to import the csv file into the newly created mservices table.

```
COPY mservices
FROM 'C:\Users\henry\Downloads\added_data\Medical\mservices.csv'
DELIMITER ','
```

CSV HEADER;

C. The highest number of diabetes cases is found in the "60+" age group across both income brackets ("Medium" and "Low"), with 406 and 490 cases respectively. This indicates the prevalence of diabetes in the older population, aligning with the understanding that the risk for such conditions increases with age.

The distribution of cases across different income brackets, particularly the high number of cases in the "Low" income bracket for the "60+" age group (490 cases), suggests that socioeconomic factors might influence the prevalence of diabetes. This could be due to various factors, including access to nutritious food, healthcare services, and health education.

When comparing diabetes and arthritis counts, there's a consistent pattern where Arthritis cases are generally higher across all segments. This might indicate that Arthritis is more widespread than diabetes across the sampled population, regardless of age and income.

Below is the query I used to determine these findings:

```
SELECT
    age_group,
    income_bracket,
    COUNT(*) FILTER (WHERE m.Diabetes = 'Yes') AS Diabetes_Count,
    COUNT(*) FILTER (WHERE m.Arthritis = 'Yes') AS Arthritis_Count
FROM (
    SELECT
```

p.patient_id,

CASE

WHEN p.age BETWEEN 18 AND 30 THEN '18-30'

WHEN p.age BETWEEN 31 AND 60 THEN '31-60'

ELSE '60+' END AS age_group,

CASE

WHEN p.income < 30000 THEN 'Low'

WHEN p.income BETWEEN 30000 AND 60000 THEN 'Medium'

ELSE 'High' END AS income_bracket

FROM patient p

) AS p

JOIN mservices m USING (patient_id)

GROUP BY age_group, income_bracket;

C1. I have attached the required csv file.

D. The mservices.csv should be acquired and refreshed monthly in the database.

D1. Patients' health conditions and the services they receive can change relatively quickly. A monthly update ensures the data reflects recent healthcare interactions. For healthcare providers and policymakers, having up-to-date information is important for making informed decisions about resource allocation, healthcare planning, and patient care strategies.

G. Sources

Office of the National Coordinator for Health Information Technology. (n.d.). *Health Information Exchange*. HealthIT.gov. Retrieved February 1, 2024, from <https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/health-information-exchange>