

**Final Project Assignment—****Pharmacy Claims**

**Chen Wang**

**ALY6030**

**Professor: Zhi He**

**03/30/2022**

**Introduction**

The assignment is a Pharmacy Claims project. We need help managers who manage third parties pay pharmacy fees and build a database from sample records. This method makes it easy for developers to set up and write SQL queries to help analysts and others analyze the claim data. However, it should be noted that, in part1, the database we set up must conform to 3NF standards, and each table should be a dimensional table or a fact table as required. In part2, we need to specify the primary and foreign keys in the table in a star schema. In part3, we need to draw the ERD of the database and send the table structure to the PBM so that they can make a decision. In part4, we need to make relevant queries and reports on the data and report the results to the company manager to provide data for project analysis.

**Part 1: Normalization**

**1: For each fact variable in your fact table, what type of fact is it? Additive, semi-additive, or non-additive?**

In the 3NF compliant database I set up, there are two facts in the fact table fact\_case\_patient\_drug. copay and insurancepaids.

copay is the fact that the patient pays part of the cost personally. Insurancepaid fact is a charge partially paid by the insurance company. Both are Additive facts. We can calculate from any dimension, whether date, patient, or drug.

**2: In your fact table, describe the grain in one sentence. What does each fact row represent?**

The grain of the fact table fact\_case\_patient\_drug can be considered a Pharmacy claim for a patient. Each row of data represents different fill\_dates and pharmacy claims of different drugs of each patient, including copays that are partly paid by individuals and Insurancepaid partly borne by insurance companies. The two cost facts in the table can be described by the dimensions patient\_information, drug\_ndc, drug\_brand,drug\_form, and fill\_date.

**Part 2: Primary and Foreign Key Setup in MySQL**

**1: What are the primary keys you designated for each of your tables? For each PK, is it a natural key or a** **surrogate key?**

In the dimension tables:

In the dim\_drug\_brand dimension table, the primary key is drug\_brand\_generic\_code, which is a natural key.

In the dim\_drug\_form dimension table, the primary key is drug\_form\_code, which is a natural key.

In the dim\_drug\_name dimension table, the primary key is drug\_ndc, which is a natural key.

In the dim\_patient\_information dimension table, the primary key is member\_id, which is a natural key.

In the fact table:

In the fact\_case\_patient\_drug fact table, the primary key is case\_id, which is a surrogate key.

**2: What are the foreign keys you designated for each of your tables? For each FK, which**

**table did you reference where that FK is listed as the PK?**

In the fact table fact\_case\_patient\_drug, I added four foreign keys, each joining four dimension tables, to ensure that the model conforms to the star schema.

The member\_id foreign key in fact\_case\_patient\_drug joins the dim\_patient\_information dimension table, and the drug\_ndc foreign key joins the dim\_drug\_name dimension table.

The drug\_brand\_generic\_code foreign key joins the dim\_drug\_brand dimension table.

The drug\_form\_code foreign key joins the dim\_drug\_form dimension table.

**3. For each FK, what did you tell MySQL to in case of deletion or update (CASCADE, SET**

**NULL, or RESTRICT)? Why did you select the option that you did for each FK?**

For each FK, I would use CASCADE to delete and update records.

Cascade automatically updates or deletes a record from the child table when updating or deleting a record from the parent table. Our database is a sample of PBM data on pharmacy claims. It is reasonable to assume that such data would be enormous and frequently deleted and updated in the real world. The deletion may be due to the user dropping out of insurance, while the update may be the addition of new users, the update of the user's claim record, and the update of the drug-related information. The relational variation provided by CASCADE can reduce our workload on the one hand and maintain data consistency and accuracy on the other hand.

**Part 3 — Entity Relationship Diagram (ERD)**

See the PDF attachment and the ERD Model in the MySQL Workbench Model file for details.

**Part 4 — Analytics and Reporting**

**Task 1 Write a SQL query that identifies the number of prescriptions grouped by drug name. Paste your output to this query in the space below here; your code should be included in your .sql file. Also answer this question: How many prescriptions were filled for the drug Ambien?**

图形用户界面, 应用程序

描述已自动生成

From the run results, we can clearly see that 5 prescriptions were filled for the drug Ambien.

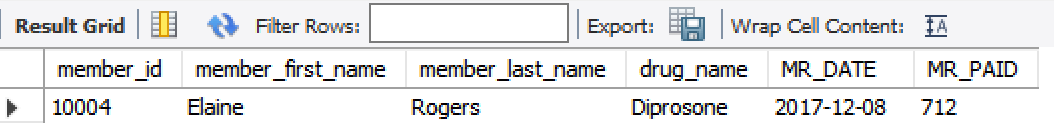
**Task 2 Write a SQL query that counts total prescriptions, counts unique (i.e. distinct) members, sums copay $$, and sums insurance paid $$, for members grouped as either ‘age 50+’ or ’ < 50’. Use case statement logic to develop this query similar to lecture 3. Paste your output in the space below here; your code should be included in your .sql file. Also answer these questions: How many unique members are over 50 years of age? How many prescriptions did they fill?**

图形用户界面, 应用程序

描述已自动生成

From the running results, we can see that there are 2 unique members older than 50. They prescribed a total of 7. Their individual payments totaled $155, and their insurance coverage totaled $2229.

**Task 3 Write a SQL query that identifies the amount paid by the insurance for the most recent prescription fill date. Use the format that we learned with SQL Window functions. Your output should be a table with member\_id, member\_first\_name, member\_last\_name, drug\_name, fill\_date (most recent), and most recent insurance paid. Paste your output in the space below here; your code should be included in your .sql file.**



**1. answer these questions: For member ID 10004, what was the drug name listed on**

**their most recent fill date?**

The most recent claim date for members of Member ID 10004 is 2017-12-08, and the drug name listed on the most recent date is Diprosone.

**2. How much did their insurance pay for that medication?**

Her insurance company covered $712 most recently for that prescription.