Introduction

Pharmacy claim databases are very useful to examine the effectiveness of drug, inappropriate medication use, to study the combinations of drugs used at random dates for treating a specific condition. Completeness of pharmacy data is very important for drug exposure and quality improvement activities. Pharmacy benefit managers (PBM) act as intermediary among pharmacies, drug manufacturers and payers, by managing formularies, negotiating discounts and processing the prescription claims.

The prescription claims for make-up members of an insurance company was provided by PBM to develop a test database and perform analysis on it for reporting purpose. As a developer, I am developing a SNOWFLAKE schema (dimensional model), a relational database and finally write some queries to extract data for reporting purpose.

Analysis

The given dataset contains patient information (id, name, age, gender), drug information (drug code, drug name, drug form code, drug brand generic code, drug form description, drug brand generic description) and payment details (fill date, copay, insurance pay).

1. **Normalization**

The initial task is to identify facts and dimensions, and to manually separate raw data into different tables that meets 3NF standards. In our dataset, there are few repeated columns fill\_date1, fill\_date2, fill\_date3 with corresponding payment columns copay1, copay2, copay3, insurance paid1, insurance paid2, insurance paid3. The 3 attributes fill\_date, copay, insurance paid have group of values. If a patient is prescribed with a medicine in future, again fill\_date, copay, insurance pay columns add next to the existing columns. This type of having data in separate columns is so confusing when we try to make any modifications for each prescription record. So, to bring the table to first normal form (1NF), all group values of each attribute are made atomic.

The given dataset is divided into one fact table, four dimension tables. Snowflake schema is the extension of star schema, where dimension tables are normalized, and data is split into different dimension tables. In snowflake schema, fact tables are surrounded by dimension tables which are again surrounded by dimension tables. The tables here are:

* “fact\_pharmacyclaim” - member\_id, drug\_ndc, fill\_date, copay, insurance\_pay
* “dimension\_patient” - member\_id, member\_first\_name, member\_last\_name, member\_birth\_date, gender
* “dimension\_drug” - drug\_ndc, drug\_name, drug\_form\_code, drug\_brand\_generic\_code
  + “dimension\_drugform” - drug\_form\_code, drug\_form\_desc
  + “dimension\_drugbrandgeneric” – drug\_brand\_generic\_code, drug\_brand\_generic\_desc

In dimension\_patient table, “age” column was removed, because as age of a patient is not same for each claim at different time periods. So, I think it is better to calculate age of a patient based on current date and date of birth.

In fact table, fact\_pharmacyclaim, copay and insurance pay both are additive facts. For example, copay can be summed across member\_id (total amount paid by each member), drug\_ndc (total amount paid for each drug) and fill\_date (amount paid for a certain period). In the same way, insurance pay can also be measured across member\_id (total insurance amount paid for each member), drug\_ndc (total insurance amount paid for each drug), fill date (total insurance amount paid over a period). Fact table contains the claim amount for each day. So, the grain would be total pay for each claim per day.

1. **Primary and Foreign key setup in MYSQL**

The below table (figure 1) shows the list of primary keys in the pharmacy claim database. With the foreign keys in fact table, it is not possible to identify each record uniquely. So, I have generated a primary key with auto increment option (surrogate key).

|  |  |  |  |
| --- | --- | --- | --- |
| Table name | Primary key | Type of PK | Identifying each record |
| fact\_pharmacyclaim | claim\_id | Surrogate | Identifies each claim record |
| dimension\_patient | member\_id | Natural | Identifies each patient/member |
| dimension\_drug | drug\_ndc | Natural | Identifies each unique drug |
| dimension\_drugform | drug\_form\_code | Natural | Identifies form of drug |
| dimension\_drugbrandgeneric | drug\_brand\_generic\_code | Natural | Identifies brand/generic of drug |

Figure 1: Primary keys and their types

The table below (figure 2) shows the list of foreign keys and their parent tables with referential integrity condition. For all tables in the list with primary and foreign key relationship, I have selected CASCADE for on update. This is because if any update (add new records/ modifications) happens in parent table, those updates should reflect in child table.

For first two parent tables in the list, if any record is deleted, I want to restrict the deletion of those records in my child table “dimension\_drug”. Because I want to keep the records of drug form, drug brand generic description in my child table. For further studies, research of medication records, to help manufacturers, it is very important to have information about medication form of a drug and its brand/generic type.

In the child table fact\_pharmacyclaim, I want to restrict the deletion of records because, I want to keep the details of drug code, member\_id in my child table. I want to know which patients were prescribed with particular medicine and what are details of those patients. This is very useful for categorizing the medicines based on age/member\_id/fill\_date to help the manufactures and also payers. So, I have selected restrict for all the tables in the list below. I believe storing pharmacy claim details are very important for future analysis. So, I never take Set null or Cascade option with on delete.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Child table | Foreign key | Parent table | On Update | On Delete |
| dimension\_drug | drug\_form\_code | dimension\_drugform | Cascade | Restrict |
| dimension\_drug | drug\_brand\_generic\_code | dimension\_drugbrandgeneric | Cascade | Restrict |
| fact\_pharmacyclaim | member\_id | dimension\_patient | Cascade | Restrict |
| fact\_pharmacyclaim | drug\_ndc | dimension\_drug | Cascade | Restrict |

Figure 2: Foreign keys and their referential integrities

1. **Entity Relationship diagram**

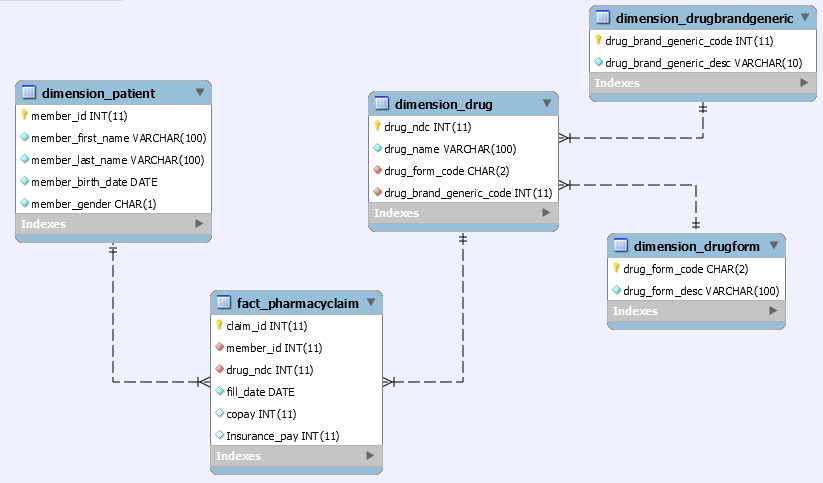
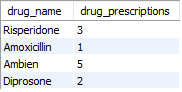


Figure 3: Snowflake schema

The snowflake schema is shown in figure 3. It is normalized form of star schema.

1. **Analytics and Reporting**

* prescriptions grouped by drug name



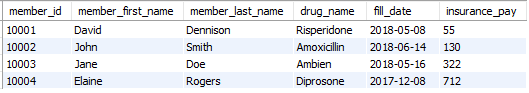
For drug name Ambien, total prescriptions are 5. It was prescribed more times compared to other drugs.

* prescription count, unique member count, total copay & insurance pay based on age group



For above 65 years age group, total unique members are only one and they fill 6 prescriptions.

* Most recent pharmacy claim records



For member\_id 10003, the drug name listed on their most recent fill date is Ambien. The insurance covered $322 for that drug.

Reference

[Julie C. Lauffenburger](https://www.ncbi.nlm.nih.gov/pubmed/?term=Lauffenburger%20JC%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), PharmD, [Akhila Balasubramanian](https://www.ncbi.nlm.nih.gov/pubmed/?term=Balasubramanian%20A%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), PhD, [Joel F. Farley](https://www.ncbi.nlm.nih.gov/pubmed/?term=Farley%20JF%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), PhD, [Cathy W. Critchlow](https://www.ncbi.nlm.nih.gov/pubmed/?term=Critchlow%20CW%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), PhD, [Cynthia D. O'Malley](https://www.ncbi.nlm.nih.gov/pubmed/?term=O%27Malley%20CD%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), PhD, [Mary T. Roth](https://www.ncbi.nlm.nih.gov/pubmed/?term=Roth%20MT%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), PharmD, MHS, [Virginia Pate](https://www.ncbi.nlm.nih.gov/pubmed/?term=Pate%20V%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), MS, and [M. Alan Brookhart](https://www.ncbi.nlm.nih.gov/pubmed/?term=Brookhart%20MA%5BAuthor%5D&cauthor=true&cauthor_uid=23696101), PhD. *Completeness of prescription information in US commercial claims databases.* Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4012425/>