ALY 6030 Data Warehousing and SQL Final Project

Anish Nitin Somaiah

[somaiah.a@husky.neu.edu](mailto:somaiah.a@husky.neu.edu)

**Part 1.**

After applying 3NF standards we have **4 dimension tables** that indicate Patient details, Drug details, Drug form & Drug brand and **1 fact table** with the Payment details as this table contain many numeric fact attributes.

1. In accordance with the first question the fact variables in our fact table fact\_paymentdetails.csv are **fill\_date, copay and insurancepaid**. Among the facts, fill\_date is semi additive, and copay & insurance are additive facts. In addition to this we have **2 foreign keys which are member\_id and drug\_ndc which are both semi additive facts.**
2. The grain in the fact table is defined by **Patient ID, Drug ID, Date and Amount Paid**. So according to these grains we can infer that each row represents what drug was consumed by which patient at which date and the amounts (insurance & copay) paid for them.

**Keynote: the other tables are dimension tables with no facts.**

**Part 2.**

1. In accordance with the first question:

* for table dim\_patientdetails our primary key is member\_id which is a natural key.
* for table dim\_drugdetails our primary key is drug\_ndc which is a natural key.
* for table dim\_drugform our primary key is drug\_form\_id which is a natural key.
* for table dim\_drugbrand our primary key is drug\_brand\_generic\_code which is a natural key.
* for table fact\_paymentdetails our primary key is payment\_id which is a surrogate key.

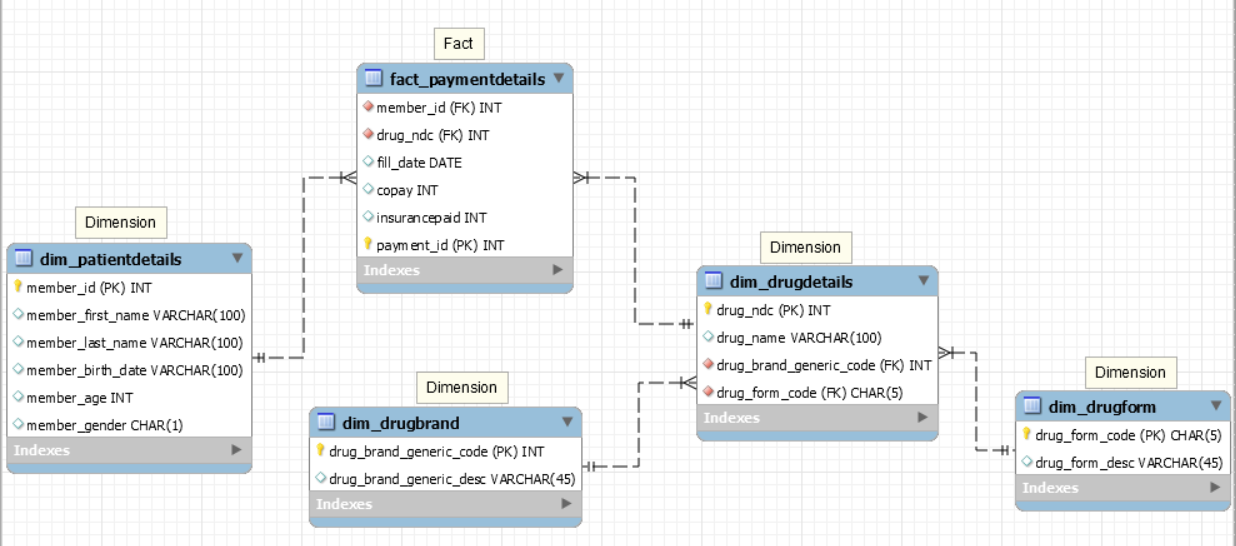
1. In accordance with the second question:

* for table fact\_paymentdetails our foreign keys are member\_id and drug\_ndc linked to dim\_patientdetails and dim\_drugdetails table respectively.
* for table dim\_drugdetails our foreign keys are drug\_form\_id and drug\_brand\_generic\_code linked to dim\_drugform and dim\_drugbrand tables.

1. In accordance to the third question:

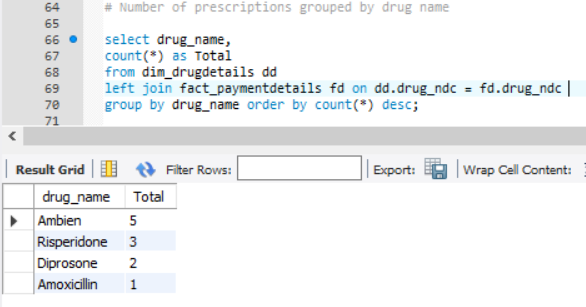
* In case of deletion for both foreign keys I have used SET NULL because data will be retained as it is and if CASCADE was used original data can be altered/removed and if RESTRICT was used no corrections can be done over the mistakes in the data and our data set will remain redundant.
* In case of updation for both foreign keys I have used RESTRICT to avoid redundancies, alteration and deletion of my existing data set which is possible with SET NULL and CASCADE functions.

**Part 3.**



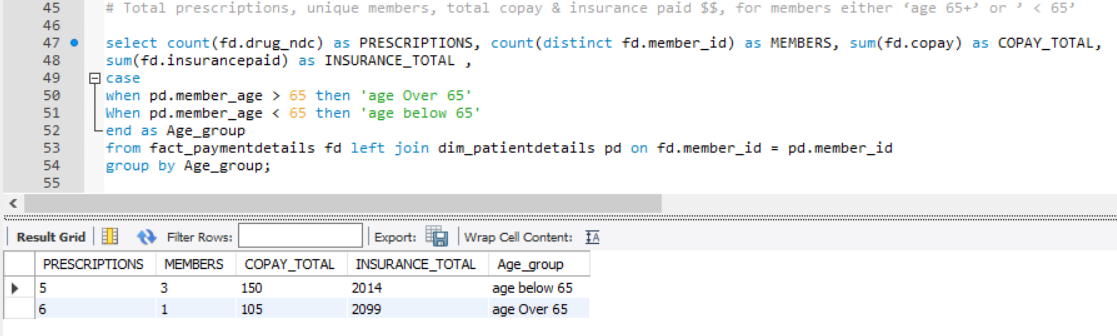
**Part 4.**

1. A SQL query that identifies the number of prescriptions **grouped by drug name**.



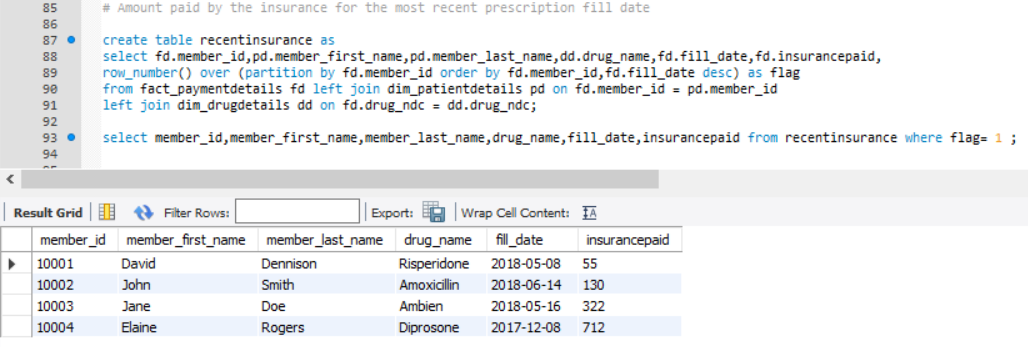
**ANS: 5 prescriptions were filled by the drug AMBIEN**

1. 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 65+’ or’ < 65’.



**ANS: There is only 1 unique member over age 65 and the member filled 6 prescriptions.**

1. A SQL query that identifies the **amount paid by the insurance** for the **most recent prescription fill date**.



**ANS: For Member ID 10003 the drug name listed in the most recent fill date is AMBIEN and insurance paid $322 for that medication.**