**Database (MySQL) manipulation for the Dataset**

**Basant Khati**

**April 2017**

**Preparing the database:**

**Step 1: Create tables and load the data**

Execute SQL scripts “references.sql” and “transactions.sql”. New database with various table will be created and data from provided dataset will be loaded into the tables.

**Step 2: Create indexes**

Create indexes for different tables based on the anticipated query fields

create index idx\_t1 on transactions(Patient\_ID,Drug\_ID);

create index idx\_t2 on ChronicIllness\_LookUp(MasterProductID,Chronicillness);

create index idx\_t3 on ATC\_LookUp(ATCLevel3Code,ATCLevel3Name,ATCLevel5Code,ATCLevel5Name);

create index idx\_t4 on Drug\_LookUp(MasterProductID,ATCLevel3Code,ATCLevel5Code);

create index idx\_t3\_5 on ATC\_LookUp(ATCLevel5Code,ATCLevel5Name);

create index idx\_t4\_5 on Drug\_LookUp(MasterProductID,ATCLevel5Code);

create index idx\_patients on patients(Patient\_ID,postcode);

**Step 3: Create views**

CREATE VIEW PatientsAndDiseasesView

AS

(SELECT t.Patient\_ID, p.postcode, p.gender, t.Prescription\_Week, t.Dispense\_Week, a.ATCLevel5Name,d.StrengthCode

FROM ATC\_LookUp a FORCE INDEX (idx\_t3\_5),

transactions t FORCE INDEX (idx\_t1),

Drug\_LookUp d FORCE INDEX (idx\_t4\_5),

Patients p FORCE INDEX(idx\_patients)

where

p.Patient\_ID = t.Patient\_ID

and

t.Drug\_ID = d.MasterProductID

and d.ATCLevel5Code = a.ATCLevel5Code);

CREATE VIEW Metformin500MGView

AS

select Patient\_ID, postcode, gender,Prescription\_Week,Dispense\_Week,ATCLevel5Name,StrengthCode from PatientsAndDiseasesView

where StrengthCode like '%500MG%' and ATCLevel5Name like 'METFORMIN%' ;

Now, database is ready for the relevant R queries.