Premise of the project:

Dr. Nice from NYU Langone is a nation renowned opioid expert interested in investigating whether patients with Opioid Use Disorder (OUD) or Opioid Poisoning (OPP) receive proper treatment after an opioid related hospital visit in 2015.

As a lead analyst of the project, you have led your team through an extensive literature review and found the following information:

1. All the opioid related diagnosis codes (ICD9/10)
2. There are 4 opioid treatments in the field, TX1 through TX4
3. Ways to identify each one of the treatment groups using procedure codes, rate codes and diagnostic codes (ICD9/10)
4. Provider IDs and names of every hospital in New York City.

You have the following datasets stored in multiple SAS libraries:

1. DX\_hospital.sas4bdat

All the diagnoses for every patient who had a hospital visit in NYC in 2015, including the following variables:

Patient\_id: unique patient ID (Stating here unique is confusing - it is not unique for this table)

New\_Tcn: unique medical visit ID (Stating here unique is confusing – same reason)

DX\_CD: Diagnostic codes

Srv\_dt\_adj: Date of diagnosis

Truadm: hospital admission date

truedisch: hospital discharge date

prov\_id: provider ID

*Other variables also exist in the file*

1. dx\_opioid.sas4bdat

The history of every opioid related diagnosis in the city, including the following variables: (seems like it’s a derived table of the previous one).

Patient\_id: unique patient ID

New\_Tcn: unique hospital visit ID (are hospital and medical visit the same?)

DX\_CD: Diagnostic codes

Srv\_dt\_adj: Date of diagnosis

prov\_id: provider ID

OUD: 1 if the diagnosis is Opioid Use Disorder

OPP: 1 if the diagnosis is Opioid Poisoning

OAE: 1 if the diagnosis is Opioid Adverse Effect

Cos\_final: categories of services, including inpatient, outpatient, mental health visits etc.

*Other variables also exist in the file*

1. tx\_opioid.sas4bdat

Patient\_id: unique patient ID

New\_Tcn: unique hospital visit ID

TX\_CD: treatment codes

Srv\_dt\_adj: Date of treatment

prov\_id: provider ID

*Other variables also exist in the file*

1. eligibility.sas4bda

Patient\_id: unique patient ID

Age: the age of the patient

Race: race of the patient

Gender: gender of the patient

*Other variables also exist in the file*

1. provider.sas4bdat

Prov\_ID: provider ID

Prov\_name: the name of the provider

Questions:

Refer to “Part 1” of the codes: (The SQL part)

1. Please explain the goal of the first SQL statement.

Adding a column to dx\_hospital table named OUDP which will be equal to 1 in case diagnostic code is an ICD10 type and it’s an Opioid Use Disorder (OUD) or Opioid Poisoning (OPP), otherwise OUDP will be equal to null.

1. What dataset(s) does the first statement use? And what is left join?

Using all the columns from the original dx\_hospital (Patient\_id, New\_Tcn, DX\_CD, Srv\_dt\_adj, Truadm, truedisch and prov\_id and others) plus the newly created OUDP column generated from collapsing dx\_opioid table to distinct dx\_cd’s and setting OUDP as 1 to columns having an ICD10 code and OUD or OPP equal 1. Generally, left join keeps all the records of the ‘left table’ and adds values from another table by matching a defined key (the ‘on’ part) – if there is no match a null value is entered. In our case the ‘left table’ is dx\_hospital and it’s joined with the dx\_opioid derived table for on a matching dx\_cd.

1. What is the function of OUDP?

OUDP is used to mark hospital visits that were entered as ICD10 code AND were related to Opioid Use Disorder (OUD) OR Opioid Poisoning (OPP).

1. You create a new dataset in the second SQL statement called dx\_hospital\_unique. What causes the duplication to begin with?

Duplication will be a result of multiple diagnostics codes at the same medical visit.

1. Based on the codes, how do you flag treatments?

(This one involves some SAS arguments)

For patients who stayed in the hospital for less than 14 days from the moment of treatment to the date of discharge, variables tx1\_14 through tx4\_14 count how many treatments they received in that session and first\_tx1\_dt through first\_tx4\_dt record the date of the early treatment. Variable Optx\_14 defines which of tx1-tx4 was first treatment to be given.

Tx\_start\_dt records when the first treatment out of the 4 was given.

In case difference between the date of the treatment and date of first treatment is less than 30 days tx1\_30 through tx4\_30 count how many treatments were given. N\_tx sums all the treatments given and if more than one was given optx\_30 defines which of tx1-tx4 was first treatment to be given.

1. Please list the libraries for the five dataset listed in the premise

Opioid\analysis, Opioid\treatment, Opioid\diagnoses, Eligibility, Cross Walks

Refer to “Part 2” of the code: (This part was pure SAS code)

1. What is the goal of the first proc tabulate statement?
2. One of the things the second proc tabulate statement produce is the total number of patients. Will it match the total number from the first proc tabulate? And why?

Seems like only male patients are taken.

1. Please explain the differences between the third proc tabulate statement and the forth proc tabulate statement.

Seems like the third statement generates a table with all OUDP values while the forth generates a table only for the ones with OUDP = 1.

Final Question:

Refer to both parts of the code:

What table does the sixth proc tabulate statement from Part 2 (line 119-129 lines weren’t numbered) create? What is optx\_30? Which lines of codes in part 1 created the necessary variables for the table?

See answer 6