INSURANCE CLAIMS DATA & ANALYTICS
HEALTHCARE DATA MINING AND DATA ANALYSIS

QINGYUE SU
MSBA，BRANDIES UNIVERSITY
qingyuesu@brandeis.edu

Table of Contents

Insurance Claims Data & Analytics

Introduction

This assignment concentrates on the analytics on the most important database in market for health, the Insurance Claim Data also known as Administrative Claim Data. These databases consist of inpatient discharge data, outpatient procedures and services data, and emergency department data. I use the available database from year 2016.

I firstly get familiar with the whole database through figuring out seven unique patients' stories by checking the data tables and related codes. Then I turn to the insurance part to compare the age distribution and spending percentage of three major insurances. Finally, I come back to the patients to examine the enormity of the health crisis related to illicit drugs and prescription opioids use/abuse/overdose.

After all these works, I get familiar with the Insurance Claim Data and are ready for deeper analysis.

Patient vignettes

In this part, I combine the tables Emergency Dept., Inpatient and Revenue Codes to find the story for each of the following patients using their UNIQ identifier:

UNIQ: 507033

Patient 507033 was admitted to Northwestern Medical Center in the first quarter of 2016. She was a young female; her age was between 25-29 at the time she was admitted. Her household zip code 054 indicates that she lived around Colchester in Vermont, where is close to the hospital's service area St. Albans. Her admission type was elective and her admission source was of non-health care facility point of origin. She subscribed to Blue Cross insurance coverage. Her main reason to visit the hospital was to deliver a baby at the hospital! She had been pregnant for 40 weeks according to DX2 code: Z3A40, and she gave a vaginal delivery to the single, live-born baby with minimal or assistance, without fetal manipulation according to DX1 code: O80. The delivery is categorized as external approach according to PX1 code: 10E0XZZ, which confirms that the process is manually assisted vaginal delivery without any instrumentation to assist in removal of the fetus. However, drainage of amniotic fluid, either via natural or artificial openings was introduced during the delivery process.

After all the delivery was smooth and successful, and the baby was in good conditions. (Congratulations to the mom!) No emergency was called, and critical access to hospital was labeled negative, though she was not discharged from the hospital on the same day. She stayed for one night in the hospital.

The delivery process incurred a bill of 3233.29 USD, of which a large portion (1002.13 USD) was paid for a semi-private 2 beds room. Another large portion of the bill was to pay for labor room, since delivery requires manual assistance. Meanwhile, there were other costs involved: 340 USD was paid for clinical diagnostic laboratory, 334.33 USD and 75.22 USD was paid for medical/surgical supplies sterile supplies and general medical/surgical supplies respectively. The rest of the cost was used to purchase pharmacy. It is concluded that room and labor are most costly during the delivery process.

UNIQ: 40436

Patient of case 40436 is an old woman. She had myocardial infarction in 2016 and was transferred from another hospital to University of Vermont Medical Center, to get better treatment.

University of Vermont Medical Center used advanced measure called Fluoroscopy of Multiple Coronary Arteries using Low Osmolar Contrast to give her a through examine. Her main problems were about heart and circulatory system, such as myocardial infarction, acute posthemorrhagic anemia, And Atherosclerotic Heart Disease. She had transient ischemic attack before so there was Presence of coronary angioplasty implant and graft in her heart. Besides heart issues, she also had those common health problems among old people, such as hyperlipidemia, diabetes mellitus, cough and asthma. In addition, I can tell from her diagnose that this woman had an unhealthy lifestyle before: She was addicted to nicotine, and was heavily dependent on insulin and aspirin.

She was given Dilation of Coronary Artery, Two Arteries with Drug-eluting Intraluminal Device, Percutaneous Approach treatment at the hospital that day. After that, she was discharged back to her home immediately. So, she only stayed in that hospital with 1 day.

All services examine and treatment cost $70275.41. But the woman has Medicare, and it includes all the charges. So, the patient didn't pay anything in this case. The main charges came from Cardiology, Medical/Surgical Supplies(implants), and Cardiac catheter lab, which reached around $42k, and accounted for 60% of the total expenditure.

UNIQ: 859382

This patient was a male in his early 30s, and he was sent to Rutland Regional Medical Center due to emergency from non-health care facility. The main reason why he was sent to medical center is that he was poisoned by heroin accidentally or unintentionally, which is classified as a kind of substance-related disorders of mental disorders. After one day's efforts of doctors, it finally turned out that all rescue measures proved ineffectual. He passed away the same day as he was sent to the hospital.

After researching the description of his diagnosis code and the doctors' procedure code on the ICD data website, I can tell the main reason why he was sent to medical center is that he was strongly addicted with drugs for a long time. Besides, after analyzing all these descriptions about his diagnosis, I can infer the process of his weakness status goes like this:

He took heroin for a long time, which can be told from the diagnosis1 and diagnosis4. Diagnosis1 tells that he was poisoned by heroin, so it may happen accidentally or unintentionally. But Diagnosis4 says that he has an opioid dependence, which means that he has already relied on drugs for a long time. This is the same as what I anticipated.

Then the heroin he took into caused the hypercapnia, which means that the CO_2 content in his blood is excessively high. This symbol contributed to some inspiration problems and neural problems, which makes the heavy burden of this brain. Besides, because of the compression of brain, he got anoxic brain damage at the same time, which made him ventricular fibrillation. This means that his heart cannot beat normally, which caused the last reason why he got so ill as cardiac arrest due to other underlying condition, and finally lead to his death.

UNIQ: 1585831

This is a 40-44 years old woman sent to Rutland Regional Medical Center ER in 2016 due to acute respiratory failure accompanied by acute pulmonary edema caused by heroin poisoning. CPR (Cardiopulmonary resuscitation) was performed and respiratory intubation (ENDO) was applied. However, infective endocarditis caused pulmonary edema, leading to hypoxemia and acidosis. She stayed in ER for one day and passed away in the following day.

The woman has history of mental disorders and substance-related disorders with Opioid dependence, benzodiazepines, and other psychoactive substance abuse. Her right lower limb has puncture wound which could be caused by herself. The woman also has metabolism disease such as diabetes, which leads to the infection of wound and cellulitis, and viral hepatitis C, which can be caused by drugs abuse.

The woman lived in Rutland with zip in 05600-05699 range, excluding Barre and Montpelier above. The woman was sent to ER (REVCODE: 450) and received CPR (REVCODE: 270) and ENDO application (respiratory intubation) (REVCODE: 410). PE (Physical Examination) test such as CT and X-ray (REVCODE: 351,320) were applied after the surgical procedure to check the edema and related cardiopulmonary problems.

Afterward, the hospital found that the woman has serious infective endocarditis and pulmonary edema, immediate laboratory bacteria culture was conducted to produce antibiotic (REVCODE: 300). During her stay at the ER (REVCODE: 200), IV solutions (REVCODE: 258) was continuously injected with Antibiotic (REVCODE: 250) and other pharmacy (REVCODE: 259). EKG and vital sign measurement were 24 hour on (REVCODE: 730).

The total cost of the service is 17093.79. The patient's major insurance resource is from MEDICAID. Her payment type is Hospital Based, regular inpatient payment.

UNIQ: 200760

The patient is a female age between 18-24, from Burlington, VT. She had Gastro- esophageal reflux disease and major depressive disorder.

In the first quarter of 2016, she encountered a motor-vehicle accident that was the first time in her life and was sent to the emergency room of University of Vermont Medical Center directly.
Her left lower limb was fractured, to be specific: she had displaced fracture of medial malleolus of left tibia and unspecified fracture of shaft of left fibula. The hospital treated her through open approach, the treatment included reposition left tibia with internal fixation device and reposition left fibula with intramed fix.

The female did not need special care or observation, after staying in hospital for 4 days, she went back home and continued her recovery.

For the patient's cost, her hospital based inpatient final bill was $49533.15. The charge covered drugs, med supplies, clinical laboratory, diagnostic radiology, anesthesiology, physical therapy and emergency room. Moreover, the female also had operations on the digestive system as an outpatient in the first quarter of 2016, which cost her $14055.80. And since the patient was at a young age, her principle payment came from commercial insurance.

UNIQ: 3692

Patient of case 3692 is a young boy. he lives in MILTON.

In the winter of 2015, he felt extremely uncomfortable, so he went to the emergency department in the University of Vermont Medical Center for help.

His illness was diagnosed as mental illness, the situation seems to be serious. So the hospital arranged him a bed to deliver further treatment.

His diagnoses include some mental illnesses such as Bipolar disorder and suicidal ideations, also include some physical hurt such as brain injury and anterior process of left calcaneus. He was also addicted to cannabis and nicotine. So, cannabis and nicotine dependence have been diagnosed. The hospital use pharmacotherapy for substance abuse treatment and detoxification services for substance abuse treatment. That's to say, the main health problems caused to this boy is due to his drug abuse.

The boy's situation was not in danger, and he wasn't sent to special care room in the hospital. However, mental illness is not an easy thing to deal with. It needs a long period to treat and take care of. He had lived in hospital for 58 days.

All the treatment spent $117895.3, it's a huge amount of cost. Fortunately, the boy had medicaid, it could pay the whole expense of his illness, which include psychiatric treatment, pharmacy cost, Laboratory - Clinical Diagnostic, Radiology - Diagnostic, and Emergency Room, and Electrocardiography. So the boy didn't need to pay anything. Among all these items, psychiatric treatment got a revenue of 106662, which is the main cost of the boy's admission to hospital.

UNIQ: 690326

The patient 690326 is a 40-44-year-old lady who comes from Milton. She went to the University of Vermont Medical Center during April to June 2016 through the elective hospital admission which means the doctor requested a bed be reserved for her. The hospital service area is Burlington, and the admission source of the patient is non-health care facility point of origin that represents she was admitted to this facility upon an order of a physician.

According to the discharge status, she is at home when the conclusion of a health care facility encounter or at the time end of a billing cycle. The lady stayed in the hospital for three days because she asked for a cosmetic surgery: a bariatric surgery. After the surgery she had some reaction of pruritus, tachycardia and muscle spasms. The services she has during her stay in the hospital including room and board (costs $5076), operating services ($25619.89), Anesthesia during the surgery ($4609.64) and CT Scan ($3273.01), etc.

The costs of these services are not being paid/partially paid by the insurance company, and she had to pay the full bill herself. The bill type is hospital based, inpatient final bill with a total charge of $43425.53.

Service and Cost Profile of Major Insurances

The top 5 MDCs paid by Medicare are 8,5,4,6 and 18, which are musculoskeletal, heart & circulatory, respiratory, digestive, and infection respectively, as shown in the pie chart. Musculoskeletal and heart & circulatory combined take almost 40% of Medicare's total spending, since Medicare beneficiaries age above 65 years old, and the elderly tend to experience musculoskeletal and heart related issues. Moreover, the largest population covered by Medicare is over 75 years old (Figure 1). The diseases related to the age group can surely explain the categories that cost Medicare a great amount of money.

The most common problems for elders are musculoskeletal related problems such as Osteoarthritis OA and low back pain. Most of these problems need constant care, and in fact most of the time it is impossible for patients to recover fully. As a result, the elderly annoyed by the diseases have to undergo some surgeries and go to physical therapy or occupational therapy often; the frequent visits to hospitals will cost a lot. Cardiopulmonary disease is another annoying problem for the elderly. Similar to musculoskeletal problems, cardiopulmonary disease takes a great amount of time to rehabilitate and recover as well. Moreover, the cost for operation of heart is considerably high and the outcome for elders is poor. Therefore, high percentage of elders undergo heart surgery will have to stay in hospital for a long time; sometimes they are trapped in hospital for the rest of their lives. Poor lung, heart, and muscular functions cause respiratory problems. In a nutshell, all these five-top MDCs are are common problems in the elderly.
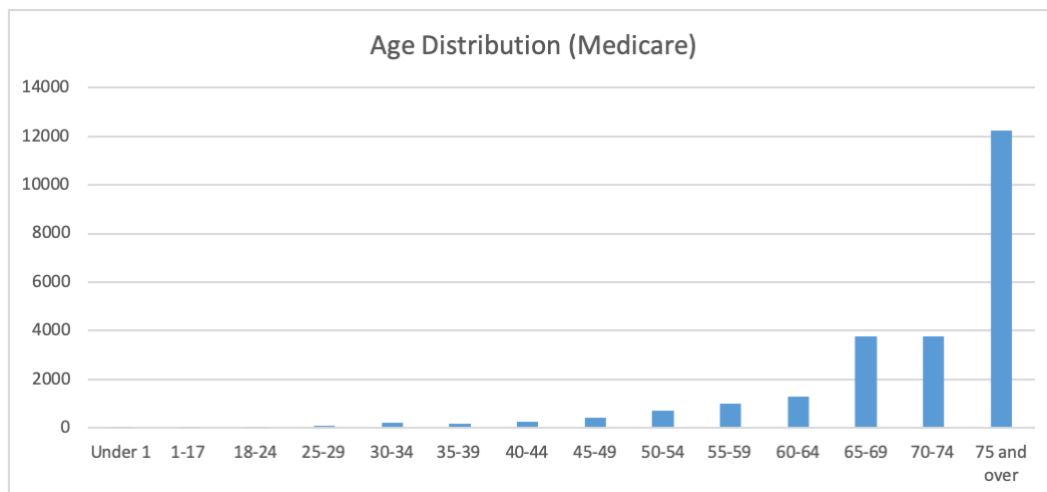

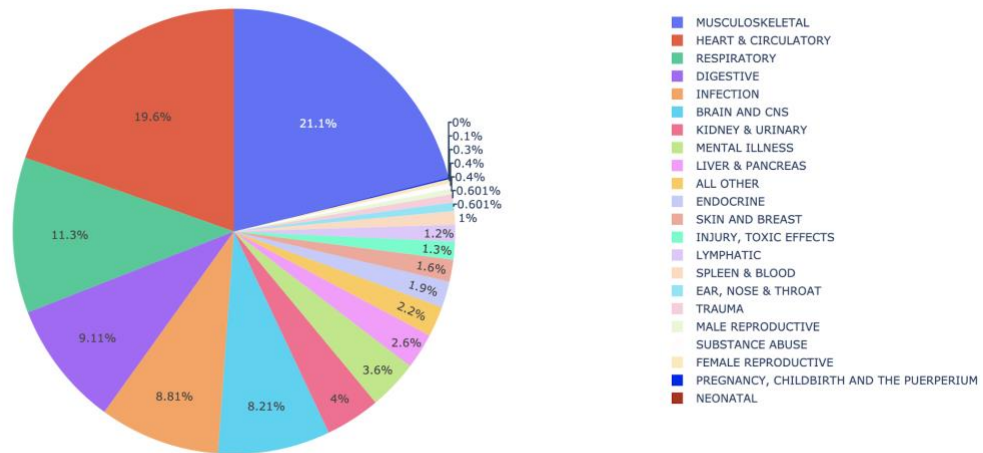Figure 1-1 Age distribution (Medicare)

Medicare MDC Spending Percentage

Figure 1-2 MDC Spending Percentage (Medicare)

The pie chart also indicates that the top 5 MDCs paid by Medicaid are 15,8, 14, 4 ,19 and 5, which are neonatal, musculoskeletal, pregnancy, childbirth and the puerperium, mental illness, and heart & circulatory respectively. The major reason that MDC 15 and 8 cost so much from Medicaid is that most of the members of Medicaid belong to low income class and other demographic groups with limited medical resources. The population includes women, children, and elders. However, most elders are covered by Medicare. Therefore, the majority of beneficiaries of Medicaid are children and women.
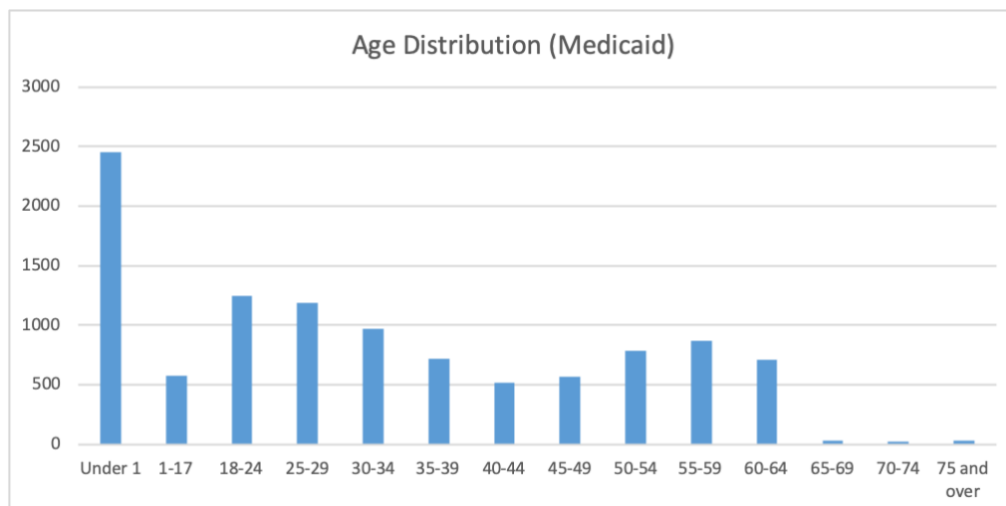

Figure 2-1 Age distribution (Medicaid)
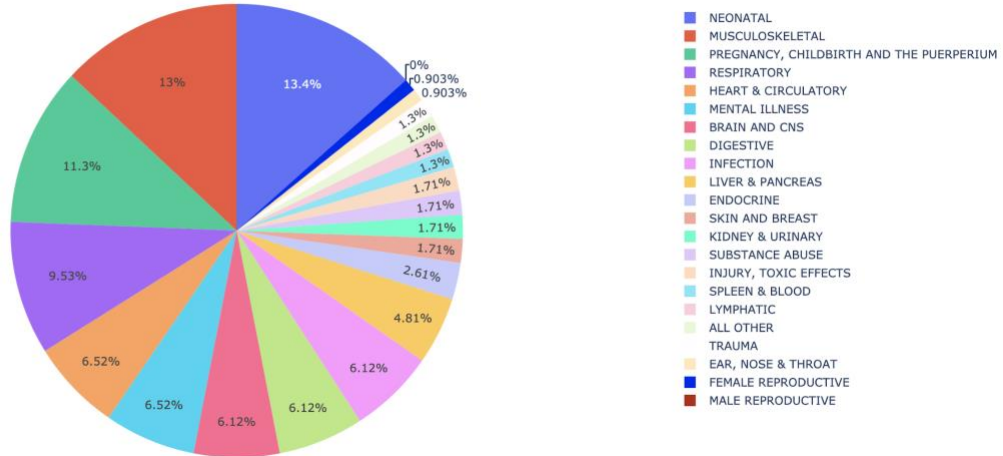
Medicaid MDC Spending Percentage

Figure 2-2 MDC Spending Percentage (Medicaid)

The gender ratio between men and women covered by Medicaid is 2 to 3. Meanwhile, the largest age group is under 1 year old. Women and children are more susceptible to diseases in pediatrics, obstetrics and gynecology categories. Also, a high percentage of low-income women have significant higher probability to suffer from mental illness compared to other demographic groups. One explanation could be that women are more vulnerable to bankruptcy and other kinds of financial difficulties owing to gender related bias in employment and business activities. The analysis matches the truth of the distribution of Medicaid.

The top 5 MDCs paid by Bluecross and Commercial Insurance Company are 8, 5,1, 6,and 14 which are musculoskeletal, heart & circulatory, brain and cns, digestive, and pregnancy, childbirth and the puerperium. The distribution of cost of MDC covered by Bluecross and Commercial Insurance Company is flatter than the previous two, since an average adult is less prone to age gender, social class related diseases.

Moreover, the two biggest cost chunks paid by commercial insurance companies are the same as the two paid by Medicare. It can be inferred that average American people suffer from musculoskeletal and heart & circulatory issues throughout their entire life. As discussed in the first paragraph, musculoskeletal and heart & circulatory related health problems take a really long time to fix. In addition, these problems are often caused by patients not holding a healthy lifestyle. For example, people who are used to sitting for hours without getting up tend to suffer from back pain. Being aware that the habit does them no good, they still find it difficult to make a change. People who do not exercise are more prone to problems such as obesity and diabetes, however, they are not motivated to go to gyms. As a result, people will suffer from these symptoms for a long time, and they are required to pay frequency visits to hospitals or therapy centers.
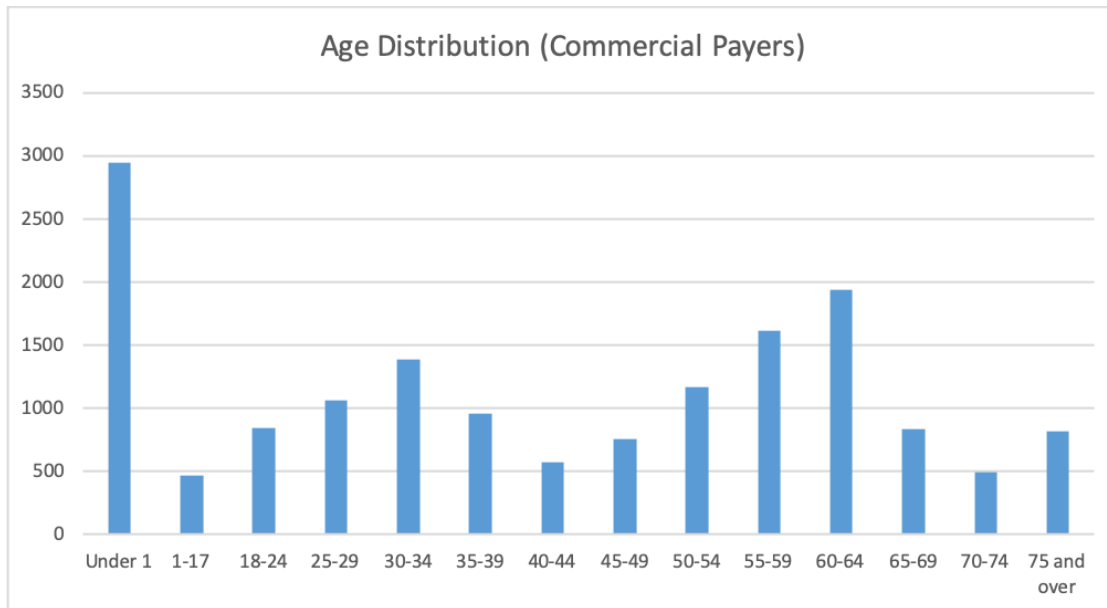
Figure 3-1 Age distribution (Commercial Payers)
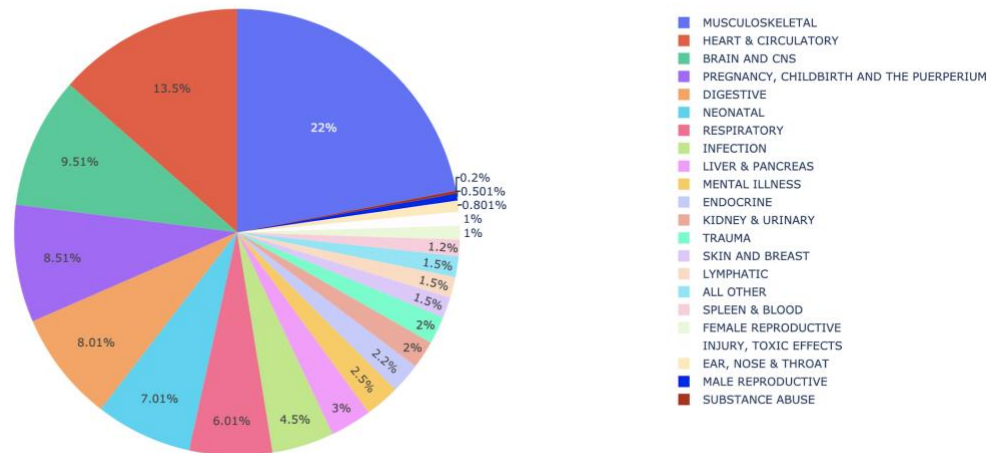

Figure 3-2 MDC Spending Percentage (Commercial Payers)

	MDC	MDC_CAT_NAME	COMMERCIAL PAYERS	MEDICAID	MEDICARE
1	1	BRAIN AND CNS	38	14	57
2	2	EYE	0	0	0
3	3	EAR, NOSE & THROAT	3	2	4
4	4	RESPIRATORY	24	22	78
5	5	HEART & CIRCULATORY	54	15	136
6	6	DIGESTIVE	32	14	63
7	7	LIVER & PANCREAS	12	11	18
8	8	MUSCULOSKELETAL	88	30	146
9	9	SKIN AND BREAST	6	4	11
10	10	ENDOCRINE	9	6	13
11	11	KIDNEY & URINARY	8	4	28
12	12	MALE REPRODUCTIVE	2	0	3
13	13	FEMALE REPRODUCTIVE	4	2	2
14	14	PREGNANCY, CHILDBIRTH AND THE PUERPERIUM	34	26	1
15	15	NEONATAL	28	31	0
16	16	SPLEEN & BLOOD	5	3	7
17	17	LYMPHATIC	6	3	8
18	18	INFECTION	18	14	61
19	19	MENTAL ILLNESS	10	15	25
20	20	SUBSTANCE ABUSE	1	4	3
21	21	INJURY, TOXIC EFFECTS	4	4	9
22	22	BURNS	0	1	1
23	23	ALL OTHER	6	3	15
24	24	TRAUMA	8	3	4
25	25	HIV	0	0	0

Examining the enormity of the health crisis related to illicit drugs and prescription opioids use/abuse/overdose

In this section, I am exploring the facts shows in the records of the emergency department. As it is known to all, illicit drugs and prescription opioids overdose is an urgent problem in the US healthcare industry. Drug overdose is usually related to emergent and severe health crises so that I focus on the emergency department to analyze drug abusers. The analysis is based on the ICD-10 code, which has a whole section T40 to T43 that related to the drug abused cases. According to the study of this part, I should be able to know about the current drug-abused situation, and this information could help policymakers and insurance companies to make adjustments on policies or supervise drug usages.

Database Description：

I filter the Emergency Dept data by only keeping the ICD-10 codes including T40xxxx, T41xxxx, T42xxxx, and T43xxxx. Since the codes can appear in any location from DX1 to DX20, I keep all the related columns as long as such code exists at least once for the patient.

Related Questions:

1. There are a total of 2151 visits in the hospital that are diagnosed as drug abusers in Emergency visits. The emergency visits have 265,862 records, which represents that near 1% of emergency visits were because of a drug overdose during 2016.

2. For the gender bias assumption, I found out that the male number of drug use/abuse is 1009, while the female number is 1141. In this case, the assumption is unreliable in our database. In fact, the drug use/abuse problem for a female is as severe as that for males.

3. The dollars spent for identified patients group by there insurance companies are as follows:

ppay(principle payment source)	description	Sum of charge($)
0	MISSING OR INVALID	219,774.24
1	MEDICARE	17,237,917.63
2	MEDICAID	6,713,846.42
4	OTHER GOVERNMENT	14,596.24
5	WORKER'S COMPENSATION	137,479.26
6	BLUE CROSS	3,342,590.94
7	COMMERCIAL INSURANCE	2,434,906.94
8	SELF PAY	264,032.70
10	NO CHARGE	66,206.52
11	CHAMPUS	24,591.95
12	HMO	285,276.69

In Question 2, BLUE CROSS and COMMERCIAL INSURANCE combine together as COMMERCIAL PAYERS. So the three insurances in Question 2 are MEDICARE, MEDICAID, and COMMERCIAL PAYERS.

| Insurance Company	Sum of Charge	Share of Total Payment
MEDICARE	17,237,917.63	56.07%
MEDICAID	6,713,846.42	21.84%
COMMERCIAL PAYERS.	5,777,498	18.79%

I can see the three major insurances consist of over 96% share of total charges, and the charge number of drug overuse is high.

4. Recently, the use of synthetic narcotics is rising but the users only know a little about these types of drugs. According to our data, 157 visits are related to these specific drug abuse cases. There are an overall 2151 visits due to drug abuse, and synthetic narcotics has made up a noticeable ratio. If I read the news, I will find that functional addiction is a common problem among those people who work under too much pressure. Students with good GPAs, lawyers, CEOs, etc. Most of the people who have functional addiction usually look okay till they were sent to the emergency room suddenly. These kinds of drugs including but not limited to Nootropics, Opioids – pain relief, relaxation, and Heroin. In the 157 samples, 114 of them are poisoning by, adverse effect of and underdosing of other synthetic narcotics, while 43 of them are poisoning by, adverse effect of and underdosing of amphetamines. There are 70 males and 87 females in the sample that shows a small difference. Therefore, everyone could be a drug abuser related to synthetic narcotics or amphetamines. The longest days a patient stayed in the hospital because of the reason is 32 days, and it is a dangerous signal to mention.

5. Name the 3 zip code regions with the highest numbers of drug use/abuse cases.

054	VT(Zip in 05400-05499 range, excluding Burlington to Saint Albans above)
057	VT(Zip in 05700-05799 range, excluding 05701)
05701	VT(RUTLAND)

3 zip code regions with the highest numbers of drug use/abuse cases
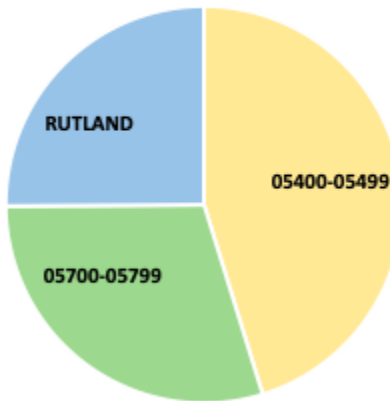
I could tell from the figure that the areas of zip codes range from 05400 to 05499 have the highest numbers of drug use/abuse cases, and then comes to the areas of zip codes from 05700 to 05799. The third one is Rutland.

6.What are the 10 most common diagnoses of drug use/abuse

ICD CODE	Diagnoses of drug use/abuse
T401X1A	Poisoning by heroin, accidental (unintentional), initial encounter
T402X5A	Adverse effect of other opioids, initial encounter
T424X2A	Poisoning by benzodiazepines, intentional self-harm, initial encounter
T424X5A	Adverse effect of benzodiazepines, initial encounter
T40605A	Adverse effect of unspecified narcotics, initial encounter
T43222A	Poisoning by selective serotonin reuptake inhibitors, intentional self-harm, initial encounter
T402X1A	Poisoning by other opioids, accidental (unintentional), initial encounter
T424X1A	Poisoning by benzodiazepines, accidental (unintentional), initial encounter
T426X5A	Adverse effect of other antiepileptic and sedative-hypnotic drugs, initial encounter

T426X2A	Poisoning by other antiepileptic and sedative-hypnotic drugs, intentional self-harm, initial encounter
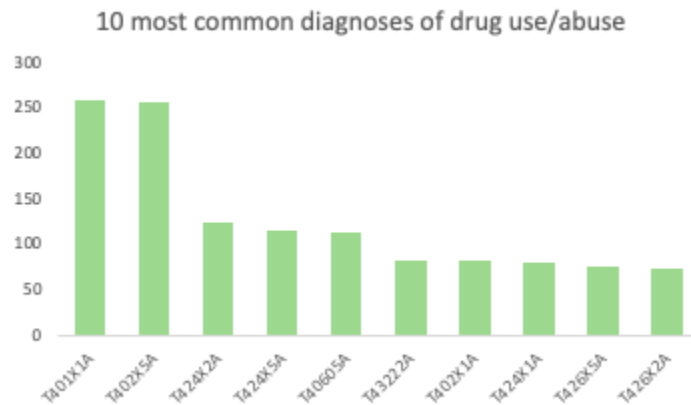
According to the results, I could tell that the most common diagnoses of drug use is Poisoning by heroin, which is caused by patients themselves. The second one is caused by other opioids, which becomes an increasingly serious problem in the US.

Appendix:

Question 2 R Markdown code:

```{r}
#(1)install packages
#install.packages("dplyr")
#install.packages("readr")
#install.packages("sandwich")
#install.packages("readxl")
#install.packages("data.table")

library(readr)
library(dplyr)
library(sandwich)
library(readxl)
library(data.table)
```

```{r}
#(2)import data
# import data - Inpatient
VTINP16_upd <-
read.csv("~/Desktop/semester1/healthcare/assignment/Assignment3/VTINP16_upd.T
XT")
#View(VTINP16_upd)
```

```
#glimpse(VTINP16_upd)
#(3)filter data
Inpatient1 <- VTINP16_upd %>% filter(PPAY %in% c(1)) %>% select(9, 77)
Inpatient1$INSURANCE <- "MEDICARE"
#View(Inpatient1)
Inpatient2 <- VTINP16_upd %>% filter(PPAY %in% c(2)) %>% select(9, 77)
Inpatient2$INSURANCE <- "MEDICAID"
#View(Inpatient2)
Inpatient3 <- VTINP16_upd %>% filter(PPAY %in% c(6, 7)) %>% select(9, 77)
Inpatient3$INSURANCE <- "COMMERCIAL PAYERS"
#View(Inpatient3)
Inpatient0 <- rbind(Inpatient1, Inpatient2, Inpatient3)
#View(Inpatient0)
#change MDC to character
MDC_explain <- read_excel("MDC_explain.xlsx")
#View(MDC_explain)
Inpatient <- merge(Inpatient0, MDC_explain, by = "MDC")
Inpatient <- na.omit(Inpatient)
View(Inpatient)
```


```{r}
#(4) draw cross table
#Set group
Inpatient <- Inpatient %>% group_by(INSURANCE, MDC, MDC_CAT_NAME) %>%
summarise(sum(CHRGS))
Inpatient$`sum(CHRGS)` <- round((Inpatient$`sum(CHRGS)`/1000000), 0)
Inpatient <- na.omit(Inpatient)
View(Inpatient)
#Establish table
charge_table <- dcast(Inpatient, MDC + MDC_CAT_NAME ~ INSURANCE)
View(charge_table)
#fwrite(charge_table, file =
"~/Desktop/semester1/healthcare/assignment/Assignment3/Inpatient.CSV")
```

Question 2 Piechart Python code:

```python
1.  import pandas as pd
2.  import plotly
3.  import plotly.graph_objects as go
4.
5.  inpatient = pd.read_csv(r'/Users/xupech/Desktop/brandeis graduate school/Academics/2019 FA
    LL/HS 256 Healthcare analytics/Third Assignment/inpatient.csv', low_memory = False)
6.  print(inpatient.columns.values)
```

```python
7.
8.   commercial_sum=inpatient['COMMERCIAL PAYERS'].sum()
9.   medicaid_sum=inpatient['MEDICAID'].sum()
10.  medicare_sum=inpatient['MEDICARE'].sum()
11.
12.  inpatient['com_percent']= round(inpatient['COMMERCIAL PAYERS']/commercial_sum,3)
13.
14.  inpatient['caid_percent']=round(inpatient['MEDICAID']/medicaid_sum, 3)
15.  inpatient['care_percent']=round(inpatient['MEDICARE']/medicare_sum, 3)
16.
17.
18.  com_pie=pd.DataFrame(inpatient.loc[inpatient['com_percent']!= 0])
19.  fig = go.Figure(data=[go.Pie(labels=com_pie['MDC_CAT_NAME'], values=com_pie['com_percent']
     )])
20.  fig.update_layout(
21.      title_text="Commercial player MDC Spending Percentage")
22.  fig.show()
23.  plotly.offline.plot(fig, filename='Commercial player MDC Spending Percentage.html')
24.
25.  caid_pie=pd.DataFrame(inpatient.loc[inpatient['caid_percent']!= 0])
26.  fig = go.Figure(data=[go.Pie(labels=com_pie['MDC_CAT_NAME'], values=com_pie['caid_percent'
     ])])
27.  fig.update_layout(
28.      title_text="Medicaid MDC Spending Percentage")
29.  fig.show()
30.  plotly.offline.plot(fig, filename='Medicaid MDC Spending Percentage.html')
31.
32.  care_pie=pd.DataFrame(inpatient.loc[inpatient['care_percent']!= 0])
33.  fig = go.Figure(data=[go.Pie(labels=com_pie['MDC_CAT_NAME'], values=com_pie['care_percent'
     ])])
34.  fig.update_layout(
35.      title_text="Medicare MDC Spending Percentage")
36.  fig.show()
37.  plotly.offline.plot(fig, filename='Medicare MDC Spending Percentage.html')
```

Question 3 R Markdown code:

```r
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(sandwich)
library(plm)
library(lmtest)
library(stargazer)
library(tidyverse)
library(knitr)
```

```
library(car)
library(readr)
library(dplyr)

VTED16 <- read.csv("~/Documents/Brandeis/Healthcare Data
Analytics/Assignment3/VTED16.TXT")
View(VTED16)

# select first three letters in DX1-DX29
as.factor(VTED16[,10:29])
VTED16_TX <- VTED16
VTED16_TX[,10:29] <- lapply(VTED16[,10:29], function(x)substr(x,1,3))
View(VTED16_TX)

# select T40xxxx, T41xxxx, T42xxxx, and T43xxxx
VTED16_TX_3 <- VTED16_TX %>% filter(rowMeans(VTED16_TX[,10:29] == "T40")>0 |
rowMeans(VTED16_TX[,10:29] == "T41") > 0|rowMeans(VTED16_TX[,10:29] ==
"T42") > 0 | rowMeans(VTED16_TX[,10:29] == "T43") > 0)
View(VTED16_TX_3)

#1 How many ED visits exactly have been diagnosed as drug user/abuser?
length(unique(VTED16_TX_3$UNIQ))
```



```{r}
#2 There is a myth that the drug use/abuse has been a male problem and that
women have much better protection measures to stay away from drug use/abuse
let alone overdoses severe enough that lead to an ED admission. Can you check
if your data supports this gender bias myth?
male = count(VTED16_TX_3 %>% filter(sex == 1))
male
female = count(VTED16_TX_3 %>% filter(sex == 2))
female
```

```{r}
#3 Tens of millions of dollars reportedly were spent on drug use related
cases that year alone. Can you find the exact dollar amount for your
identified patients in this question? Of the three insurances in Question 2,
what was share of each of the total payments?
VTED16_TX_3_spend = tapply(VTED16_TX_3$CHRGS,VTED16_TX_3$PPAY ,sum)
VTED16_TX_3_spend
```

```r
total_spend = sum(VTED16_TX_3$CHRGS)
total_spend
share_medicare = VTED16_TX_3_spend[2]/total_spend
share_medicaid = VTED16_TX_3_spend[3]/total_spend
share_commercial_payers=
(VTED16_TX_3_spend[6]+VTED16_TX_3_spend[7])/total_spend
share_insurance = data.frame(variable =
c('share_medicare','share_medicaid','share_commercial_payers'),
      value = c(share_medicare,share_medicaid,share_commercial_payers))
```


```{r}
# 4
ed_drug = merge(VTED16_TX_3, VTED16, all.x = TRUE, by =  'UNIQ')
ed_drug = data.table(ed_drug)
data_drug_abuse = ed_drug[, -c(2:70)]

er_visits = ed_drug[, .N]
row_index = vector()
for (i in 1:nrow(data_drug_abuse)){
  if (any(startsWith(as.character(data_drug_abuse[i,11:30]),"T404"))|
      any(startsWith(as.character(data_drug_abuse[i,11:30]),"T4362"))) {
    row_index = c(row_index,i)
  }
}
sn = data_drug_abuse[row_index,]
View(sn)

sn = data.table(sn)
male = sn[sex.y == 1]
View(sn[max(pdays.y)])
```


```{r}
#5 Name the 3 zip code regions with the highest numbers of drug use/abuse
cases.
as.factor(VTED16_TX_3$TXTZIP)
VTED16_TX_3_ZIP <- VTED16_TX_3 %>% group_by(TXTZIP) %>% summarise(count =
n()) %>% arrange(-count)
VTED16_TX_3_ZIP[1:3,]
```


```{r}
```

```
#6 What are the 10 most common diagnoses of drug use/abuse?
merge_table <- merge(VTED16, VTED16_TX_3, by = "UNIQ")
merge_table_new <- merge_table[,11:30]
View(merge_table_new)
# count number
code_frequency <- table(unlist(merge_table_new))
code_frequency <- data.frame(code_frequency)
View(code_frequency)
# select useful rows
as.factor(code_frequency$Var1)
code_frequency_t <- code_frequency %>%
filter((substring(code_frequency$Var1,1,3) == "T40") > 0 |
(substring(code_frequency$Var1,1,3) == "T41") > 0 |
(substring(code_frequency$Var1,1,3) == "T42") > 0 |
(substring(code_frequency$Var1,1,3) == "T43") > 0) %>% arrange(-Freq)
head(code_frequency_t, 10)
```
```